

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:24:24 ; Search time 25.25 seconds
(without alignments)
290.327 Million cell updates/sec

Title: US-09-215-435-225

Perfect score: 1224

Sequence: 1 MGWTMRLVTAALLGLLMVV.....APRRASEPKHKNAQIAAC 227

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	235.5	19.2	187	1	PEBP_RAT
2	234.5	19.2	221	1	PBPH_CAEEL
3	231	18.9	152	1	D1_ONCVO
4	223	18.2	134	1	D3_ONCVO
5	222	18.1	197	1	OV16_ONCVO
6	219	17.9	114	1	D2_ONCVO
7	219	17.9	187	1	PEBP_MOUSE
8	218.5	17.9	186	1	PEBP_BOVIN
9	206	16.8	186	1	PEBP_MACFA
10	205	16.7	186	1	PEBP_HUMAN
11	169.5	13.8	262	1	TE26_TOXCA
12	168	13.7	210	1	OBA5_DROME
13	148	12.1	181	1	CEN_ANTMA
14	137.5	11.2	190	1	PBP_PLAFA
15	133	10.9	201	1	YL79_YEAST
16	122.5	10.0	219	1	DKAL_YEAST
17	113.5	9.3	197	1	YJ10_MYCTU
18	113	9.2	151	1	Y273_METH
19	106	8.7	171	1	YC50_AQUAE
20	105	8.6	198	1	YC69_PYRHO
21	96.5	7.9	150	1	Y877_CHLUPN
22	95	7.8	183	1	YBCL_ECOLI
23	95	7.8	201	1	YJ11_MYCTU
24	92	7.5	179	1	Y502_STRCO
25	91.5	7.5	551	1	SYE_ARCFU
26	89.5	7.3	176	1	YL40_MYCTU
27	89.5	7.3	1705	1	PTPO_MOUSE
28	86.5	7.1	150	1	Y736_CHLTR
29	85	6.9	595	1	SNX3_HUMAN
30	80.5	6.6	287	1	Y98_ARCFU
31	79	6.5	265	1	HL_PEA
32	79	6.5	602	1	TX15_MOUSE
33	77.5	6.3	490	1	ACM4_CHICK

34	76	6.2	748	1	MEPA_RAT
35	76	6.2	1102	1	P11G_PIG
36	76	6.2	2607	1	BACB_BACLI
37	75.5	6.2	1130	1	REPT_MOUSE
38	75	6.1	466	1	ACM2_CHICK
39	75	6.1	1041	1	EGT2_YEAST
40	74.5	6.1	630	1	INLB_LISMO
41	74	6.0	659	1	GLGX_HAEIN
42	74	6.0	1004	1	YD83_SCHPO
43	73.5	6.0	240	1	GTXH_MOUSE
44	73.5	6.0	382	1	YMF4_CAEEL
45	73.5	6.0	397	1	TBXT_CHICK

ALIGNMENTS

```
RESULT 1
PEBP_RAT
ID PEBP_RAT STANDARD; PRT; 187 AA.
AC P31044; P31045;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN (PEBP) (23 KDA MORPHINE-
DE BINDING PROTEIN) (P23K).
GN PEBP OR PBP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE=91042640; PubMed=1978248;
RA Grandy J.M., Hanneman E., Bunzow J., Shih M., Machida C.A.,
RA Bidlack J.M., Civelli O.;
RT "Purification, cloning, and tissue distribution of a 23-kDa rat
RT protein isolated by morphine affinity chromatography.";
RL Mol. Endocrinol. 4:1370-1376(1990).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=HIPPOCAMPUS;
RA Tohdoh N., Tojo S., Agui H., Ojika K.;
RA Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=EPIDIDYMIS, AND LIVER;
RX MEDLINE=94311839; PubMed=8037677;
RA Perry A.C.F., Hall L., Bell A.E., Jones R.;
RT "Sequence analysis of a mammalian phospholipid-binding protein from
RT testis and epididymis and its distribution between spermatozoa and
RT extracellular secretions.";
RL Biochem. J. 301:235-242(1994).
[4]
RN SEQUENCE OF 40-56 AND 93-112.
RC TISSUE=SPERM;
RX MEDLINE=92031654; PubMed=1932083;
RA Jones R., Hall L.;
RT "A 23 kDa protein from rat sperm plasma membranes shows sequence
RT similarity and phospholipid binding properties to a bovine brain
RT cytosolic protein.";
RL Biochim. Biophys. Acta 1080:78-82(1991).
CC -I- FUNCTION: BINDS PHOSPHATIDYLETHANOLAMINE. HAS LOWER AFFINITY FOR
CC PHOSPHATIDYLINOSITOL AND PHOSPHATIDYLCHOLINE.
CC -I- SUBUNIT: HAS A TENDENCY TO FORM DIMERS BY DISULFIDE CROSS-LINKING.
CC -I- SUBCELLULAR LOCATION: THERE ARE TWO FORMS: A CYTOPLASMIC FORM AND
CC A MEMBRANE-BOUND FORM.
CC -I- TISSUE SPECIFICITY: MAJOR COMPONENT OF EPIDIDYMAL SECRETIONS AND
CC SPERM PLASMA MEMBRANES. IT IS PRESENT IN CYTOSOLS FROM A VARIETY
CC OF OTHER TISSUES.
CC -I- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
CC PROTEIN FAMILY.
-----
```

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X75253; CAA53032.1; -
 KW EMBL: X75254; CAA53033.1; -
 SQ EMBL: X71873; CAA50708.1; -
 DR PIR: A36126; A36126.
 DR PIR: S37555; S37555.
 DR PIR: S18358; S18358.
 DR INTERPRO: IPR001858; -
 DR PFAM: PF01161; PBP; 1.
 DR PROSITE: PS01220; PBP; 1.
 KW Lipid-binding ?1
 FT MOD_RES ?1 ?1 BLOCKED.
 FT CONFLICT 49 R -> G (IN REF. 4).
 FT CONFLICT 54 SW -> TA (IN REF. 4).
 SQ SEQUENCE 187 AA; 20801 MW; F2BF053FE34B8056 CRC64;

Query Match 19.2%; Score 235.5; DB 1; Length 187;
 Best Local Similarity 36.3%; Pred. No. 1.2e-15;
 Matches 57; Conservative 20; Mismatches 60; Indels 19; Gaps 5;
 QY 52 ELGNIGCKVDPCCNNYRQKITSMEPIVKFPGAVDGYIILVMVDPDAPSRAEPRQFWR 111
 DB 36 ELGKV---LTPQVNNRPSSTISW-----DGLDPGKLYTLVLTDPDAPSKDPKPREWH 85
 QY 112 HMLVTDIKGADLKKKGKIQGQELSAIYQAPSPAHSGFHYOFFVYVLOGKVIS---LLPK 167
 DB 86 HFLVNNKGNDISSTGV---LSYVSGGPKDGLHRYVNLVYEQOPLNCDEPILSNK 141
 QY 168 ENKTRGSKMDFRNLHRLGPEASTQFTMTQNYQDS 203
 DB 142 SGDNRGRKRVESFRKKYHLGAPVAGTCFAE-WDDS 176

RESULT 2
 ID BPH CAEEL STANDARD; PRT; 221 AA.
 AC O16264;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN HOMOLOG F40A3.3.
 GN F40A3.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 CC Rhabdicolidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Geisel C., Bradshaw H., Keppler D.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -|- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
 CC PROTEIN FAMILY.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AF016423; AAB65322.1; -
 DR WORMPEP: F40A3.3; CE10146.
 DR INTERPRO: IPR001858; -
 DR PFAM: PF01161; PBP; 1.

DR PROSITE: PS01220; PBP; 1.
 KW Hypothetical protein: Lipid-binding.
 SQ SEQUENCE 221 AA; 24143 MW; 4828CC749D8B32AC CRC64;

Query Match 19.2%; Score 234.5; DB 1; Length 221;
 Best Local Similarity 33.1%; Pred. No. 1.8e-15;
 Matches 58; Conservative 16; Mismatches 60; Indels 41; Gaps 6;

QY 59 KVVDP--CNNYRQKITSMEPIVKFPGAVD-----CATYIL 92
 DB 46 EVIDPVLASNPSPKVS-----VKFNSGVEANLGNLWLTPTQVKDTPEVKWDAEPGALLTL 100
 QY 93 VNVDPDAPSRAEPRQFWRHMLVTDIKGADLKKKGKIQGQELSAIYQAPSPAHSGFHYQF 152
 DB 101 IKTDPDAPSRAEPRQFWRHMLVNLPGNDIAK----GDTLSEYIGAGPPPKTGLHRYV 156
 QY 153 FVYVLOGKVIS-----LLPKENKTRGSKMDFRNLHRLGPEASTQFTMTQNYQD 202
 DB 157 LIYKQSGRIEDAEHGRLLTNTSGDKRGKNAADFVAKHKLGAAPVFGNLFQAE-YDD 210

RESULT 3
 ID DL ONCVO STANDARD; PRT; 152 AA.
 AC P54186;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE D1 PROTEIN (FRAGMENT).
 GN D1.
 OS Onchocerca volvulus.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 CC Onchocercidae; Onchocerca.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97045813; PubMed-8890735;
 RA Ertmann K.D., Gallin M.Y.;
 RT "Onchocerca volvulus: identification of cDNAs encoding a putative phosphatidyl-ethanolamine-binding protein and a putative partially processed mRNA precursor."
 RT processed mRNA precursor."
 RL Gene 174:203-207 (1996).
 CC -|- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
 CC PROTEIN FAMILY.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X87991; CAA61244.1; -
 DR INTERPRO: IPR001858; -
 DR PFAM: PF01161; PBP; 1.
 DR PROSITE: PS01220; PBP; 1.
 FT NON_TER 1
 SQ SEQUENCE 152 AA; 16767 MW; E3B68C2E3E3B295D CRC64;

Query Match 18.9%; Score 231; DB 1; Length 152;
 Best Local Similarity 37.0%; Pred. No. 2.5e-15;
 Matches 54; Conservative 18; Mismatches 58; Indels 16; Gaps 5;

QY 55 NIGCKVVP-DCNNYRQKITSMEPIVKFPGAVDGYIILVMVDPDAPSRAEPRQFWRH 113
 DB 20 NLGNELTPTQVKQPTKV-SW-----DAEPGALYTLVMTDPDAPSRAEPRQFWRH 70
 QY 114 LVTDIKGADLKKKGKIQGQELSAIYQAPSPAHSGFHYOFFVYVLOGKVISLLPKENKTRG 173
 DB 71 LIINISQNVSSGTV-----LSDYIGSGPKGTGLHRYVFLVYKQPGSITD--TQHGNGRP 124

Query Match 18.1%; Score 222; DB 1; Length 197;
Best Local Similarity 36.3%; Pred. No. 2.5e-14;
Matches 53; Conservative 18; Mismatches 59; Indels 16; Gaps 5;

QY 55 NTGCKVVP-DCNNYRQKITSWMEPIVKFPFGAVDGATYILVMVDPDAPSRAEPORFWRHW 113
DB 65 NLGNELTPQVKNQPTKV-SW-----DAEPGALYTLVMTDAPSRKNPVFREWHHW 115
QY 114 LVTDIKGADLKKKGIOGQELSAYQAPSPAHSGFHRFYFFVYLOGEKVISLLPENKTRG 173
DB 116 LIINISGQNVSSGTV-----LSDYIGSGPKGTGLHRYVFLYKPGSITD--TQHGNGRR 169
QY 174 SWKMDRFLNRFHLGPEASTQFMFTQN 199
DB 170 NEKVMDFANKHHLGNPVAGNFFQAKH 195

RESULT 6
D2_ONCV0
ID D2_ONCV0 STANDARD; PRT; 114 AA.
AC P54187;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE D2 PROTEIN (FRAGMENT).
OS D2.
GN Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
RN [1]
RX MEDLINE-97045813; PubMed-8890735;

QY 174 SWKMDRFLNRFHLGPEASTQFMFTQN 199
DB 125 NEKVMDFANKHHLGNPVAGNFFQAKH 150

RESULT 4
D3_ONCV0
ID D3_ONCV0 STANDARD; PRT; 134 AA.
AC P54188;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE D3 PROTEIN (FRAGMENT).
OS Onchocerca volvulus.
GN Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
RN [1]
RX MEDLINE-97045813; PubMed-8890735;
RT "Onchocerca volvulus: identification of cDNAs encoding a putative phosphatidyl-ethanolamine-binding protein and a putative partially processed mRNA precursor.";
RT processed mRNA precursor.";
RL Gene 174:203-207(1996).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC or send an email to license@isb-sib.ch).
CC EMBL; X87989; CA61242.1; -;
DR INTERPRO: IPR001858; -;
DR PFAM: PF01161; PBP; 1.
DR PROSITE: PS01220; PBP; 1.
FT SIGNAL 1 16
FT CHAIN 17 197
FT CARBOHYD 56 56
FT CARBOHYD 61 61
FT CARBOHYD 119 119
FT CARBOHYD 124 124
SQ SEQUENCE 134 AA; 14880 MW; AC98F5C500A251P2 CRC64;

Query Match 18.2%; Score 223; DB 1; Length 134;
Best Local Similarity 36.3%; Pred. No. 1.3e-14;
Matches 53; Conservative 18; Mismatches 59; Indels 16; Gaps 5;

QY 55 NTGCKVVP-DCNNYRQKITSWMEPIVKFPFGAVDGATYILVMVDPDAPSRAEPORFWRHW 113
DB 2 NLGNELTPQVKNQPTKV-SW-----DAEPGALYTLVMTDAPSRKNPVFREWHHW 52
QY 114 LVTDIKGADLKKKGIOGQELSAYQAPSPAHSGFHRFYFFVYLOGEKVISLLPENKTRG 173
DB 53 LIINISGQNVSSGTV-----LSDYIGSGPKGTGLHRYVFLYKPGSITD--TQHGNGRR 106
QY 174 SWKMDRFLNRFHLGPEASTQFMFTQN 199
DB 107 NEKVMDFANKHHLGNPVAGNFFQAKH 132

RESULT 5
OV16_ONCV0
ID OV16_ONCV0 STANDARD; PRT; 197 AA.
AC P31729;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE OV-16 ANTIGEN PRECURSOR.
GN OV16.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
RN [1]
RX MEDLINE-97045813; PubMed-8890735;

Query Match 18.1%; Score 222; DB 1; Length 197;
Best Local Similarity 36.3%; Pred. No. 2.5e-14;
Matches 53; Conservative 18; Mismatches 59; Indels 16; Gaps 5;

QY 55 NTGCKVVP-DCNNYRQKITSWMEPIVKFPFGAVDGATYILVMVDPDAPSRAEPORFWRHW 113
DB 65 NLGNELTPQVKNQPTKV-SW-----DAEPGALYTLVMTDAPSRKNPVFREWHHW 115
QY 114 LVTDIKGADLKKKGIOGQELSAYQAPSPAHSGFHRFYFFVYLOGEKVISLLPENKTRG 173
DB 116 LIINISGQNVSSGTV-----LSDYIGSGPKGTGLHRYVFLYKPGSITD--TQHGNGRR 169
QY 174 SWKMDRFLNRFHLGPEASTQFMFTQN 199
DB 170 NEKVMDFANKHHLGNPVAGNFFQAKH 195

RESULT 6
D2_ONCV0
ID D2_ONCV0 STANDARD; PRT; 114 AA.
AC P54187;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE D2 PROTEIN (FRAGMENT).
OS D2.
GN Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
RN [1]
RX MEDLINE-97045813; PubMed-8890735;

QY 174 SWKMDRFLNRFHLGPEASTQFMFTQN 199
DB 125 NEKVMDFANKHHLGNPVAGNFFQAKH 150

RESULT 4
D3_ONCV0
ID D3_ONCV0 STANDARD; PRT; 134 AA.
AC P54188;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE D3 PROTEIN (FRAGMENT).
OS Onchocerca volvulus.
GN Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
RN [1]
RX MEDLINE-97045813; PubMed-8890735;
RT "Onchocerca volutus: identification of cDNAs encoding a putative phosphatidyl-ethanolamine-binding protein and a putative partially processed mRNA precursor.";
RT Gene 174:203-207(1996).
RL Gene 174:203-207(1996).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN FAMILY.
CC EMBL; X87989; CA61242.1; -.
DR INTERPRO: IPR001858; -.
DR PFAM: PF01161; PBP; 1.
DR PROSITE: PS01220; PBP; 1.
FT SIGNAL 1 16
FT CHAIN 17 197
FT CARBOHYD 56 56
FT CARBOHYD 61 61
FT CARBOHYD 119 119
FT CARBOHYD 124 124
SQ SEQUENCE 134 AA; 14880 MW; AC98F5C500A251P2 CRC64;

Query Match 18.2%; Score 223; DB 1; Length 134;
Best Local Similarity 36.3%; Pred. No. 1.3e-14;
Matches 53; Conservative 18; Mismatches 59; Indels 16; Gaps 5;

QY 55 NTGCKVVP-DCNNYRQKITSWMEPIVKFPFGAVDGATYILVMVDPDAPSRAEPORFWRHW 113
DB 2 NLGNELTPQVKNQPTKV-SW-----DAEPGALYTLVMTDAPSRKNPVFREWHHW 52
QY 114 LVTDIKGADLKKKGIOGQELSAYQAPSPAHSGFHRFYFFVYLOGEKVISLLPENKTRG 173
DB 53 LIINISGQNVSSGTV-----LSDYIGSGPKGTGLHRYVFLYKPGSITD--TQHGNGRR 106
QY 174 SWKMDRFLNRFHLGPEASTQFMFTQN 199
DB 107 NEKVMDFANKHHLGNPVAGNFFQAKH 132

RESULT 5
OV16_ONCV0
ID OV16_ONCV0 STANDARD; PRT; 197 AA.
AC P31729;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE OV-16 ANTIGEN PRECURSOR.
GN OV16.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
RN [1]
RX MEDLINE-97045813; PubMed-8890735;

```
RA Erttmann K.D., Gallin M.Y.;
RT "Onchocerca volvulus: identification of cDNAs encoding a putative
RT phosphatidyl-ethanolamine-binding protein and a putative partially
RT processed mRNA precursor.";
RL Gene 174:203-207(1996).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
CC PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X87990; CAA61243.1; -
CC DR INTERPRO: IPR001858; -
CC DR PFAM: PF01161; PBP; 1.
CC DR PROSITE: PS01220; PBP; 1.
CC FT NON_TER 1
CC SQ SEQUENCE 114 AA; 12786 MW; 558494D94E2FCE72 CRC64;

Query Match 17.9%; Score 219; DB 1; Length 114;
Best Local Similarity 39.7%; Pred. No. 2.5e-14;
Matches 46; Conservative 14; Mismatches 50; Indels 6; Gaps 2;

QY 84 AVDGATYILVMVDPDAPSRAPRFRHVLVTDIKGADLKKGKIQGELSAYQAPSPPA 143
DB 3 AEPGALYTLVMTDPDVPKRPVFRWHHLIINISGVSSGIV---LSDYIGSGPPK 58
QY 144 HSGFHYRYFFVYLOEGKVYISLLPKENTRGSKWMDRFLNRFHLGEPASTQFMQTN 199
DB 59 GTGLHRYVFLVKPGSITD--TQGGNRRNFKVMDFANKHHLGNPVAGNFFQAKH 112

RESULT 7
PEBP_MOUSE STANDARD; PRT; 187 AA.
ID PEBP_MOUSE
AC P70296;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN (PEBP).
GN PEBP OR PBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RP SEQUENCE FROM N.A.
RA Lin B., Frischauf A.-M.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS PHOSPHATIDYLETHANOLAMINE. HAS LOWER AFFINITY FOR
CC PHOSPHATIDYLINOSITOL AND PHOSPHATIDYLCHOLINE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
CC PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U43206; AAB06983.1; -
CC DR INTERPRO: IPR001858; -
CC DR PFAM: PF01161; PBP; 1.
CC DR PROSITE: PS01220; PBP; 1.
CC KW Lipid-binding.
CC SQ SEQUENCE 187 AA; 20860 MW; E51DD693A0A8B8BD CRC64;
```

```
Query Match 17.9%; Score 219; DB 1; Length 187;
Best Local Similarity 34.4%; Pred. No. 4.6e-14;
Matches 52; Conservative 22; Mismatches 59; Indels 18; Gaps 4;

QY 52 ELGNIGCKVVPDCNNYRQKITSMPEIVKFKGAVDGTATYILVMVDPDAPSRAPRQRFWR 111
DB 36 ELGKV---LTPTQVMNRPSISW-----DGLDPGKLYTLVLTDPDAPSRKDPKRFWRH 85
QY 112 HMLVTDIKGADLKKGKIQGELSAYQAPSPPAHSGFHYRYFFVYLOEGKVIS----LLPK 167
DB 86 HELVNMKMDISSGTV-----LSDYVGSPPSGTSLHRYVWLVLVYEQEPLSCDEPILSNK 141
QY 168 ENKTRGSKWMDRFLNRFHLGEPASTQFMQ 198
DB 142 SGDNRGKFKVETFRKKYINLGAPVAGTCYQAE 172
```

```
RESULT 8
PEBP_BOVIN STANDARD; PRT; 186 AA.
ID PEBP_BOVIN
AC P13696;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN (PEBP) (BASIC CYTOSOLIC 21
DE KDA PROTEIN).
GN PEBP OR PBP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
[1]
RP SEQUENCE.
RA Schoentgen F., Saccoccio F., Jolles J., Bernier I., Jolles P.;
RT "Complete amino acid sequence of a basic 21-kDa protein from bovine
RT brain cytosol.";
RL Eur. J. Biochem. 166:333-338(1987).
[2]
RN X-RAY CRYSTALLOGRAPHY (1.84 ANGSTROMS).
RP MEDLINE:98455508; PubMed-9782057;
RA Serre L., Vallée B., Bureau N., Schoentgen F., Zelwer C.;
RT "Crystal structure of the phosphatidylethanolamine-binding protein
RT from bovine brain: a novel structural class of phospholipid-binding
RT proteins."
RL Structure 6:1255-1265(1998).
CC -1- FUNCTION: BINDS PHOSPHATIDYLETHANOLAMINE. HAS LOWER AFFINITY FOR
CC PHOSPHATIDYLINOSITOL AND PHOSPHATIDYLCHOLINE.
CC -1- SUBUNIT: HAS A TENDENCY TO FORM DIMERS BY DISULFIDE CROSS-LINKING.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
CC PROTEIN FAMILY.
CC PIR: S00056; S00056.
CC PDB: 1A44; 13-JAN-99.
CC PDB: 1B7A; 27-JAN-99.
CC INTERPRO: IPR001858; -
CC PFAM: PF01161; PBP; 1.
CC PROSITE: PS01220; PBP; 1.
CC KW Lipid-binding; 3D-structure.
CC INIT_MET 0
CC FT SEQUENCE 186 AA; 20854 MW; 19F17089D2F06DF4 CRC64;
```

```
Query Match 17.9%; Score 218.5; DB 1; Length 186;
Best Local Similarity 33.3%; Pred. No. 5.1e-14;
Matches 53; Conservative 24; Mismatches 65; Indels 17; Gaps 4;

QY 46 LEVFI--PELGNIGCKVVPDCNNYRQKITSMPEIVKFKGAVDGTATYILVMVDPDAPSR 103
DB 24 LOVKYGGAEVDGLKVLTPQVKNRPTSITW-----DGLDPGKLYTLVLTDPDAPSRK 76
```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D16111; BAA03684.1; -

CC EMBL; X72522; CAA53031.1; -

CC EMBL; X85033; CAA59404.1; -

CC EMBL; S76773; AADI4234.1; -

CC PDB; 1BD9; 16-SEP-98.

CC PDB; 1BEH; 16-SEP-98.

CC SWISS-2DPAGE; P30086; HUMAN.

CC INTERPRO; IPR001858; -

CC PFAM; PF01161; PBP; 1.

CC PROSITE; PS01220; PBP; 1.

CC Lipid-binding; 3D-structure.

CC INIT_MET 0 0

CC VARIANT 8 8 S -> N.

CC CONFLICT 7 7 /FTID=VAR_006048.

CC W -> K (IN REF. 2).

CC SEQUENCE 186 AA; 20925 MW; F18F2AC30747C78B CRC64;

Query Match 16.7%; Score 205; DB 1; Length 186;

Best Local Similarity 33.1%; Pred. No. 1e-12;

Matches 50; Conservative 21; Mismatches 62; Indels 18; Gaps 4;

QY 52 ELGNTGCKVPCNNYRKITSWMEPIVKFPGAVDGYILVWYDPDAPSPRAEPQRQFWR 111

DB 35 ELGKV---LTPTQVKNRPTISW-----DGLDSGLKYLTLVLDTPDAPSRKDPKYREWH 84

QY 112 HWLVTDIKGADLKKGIQQLSAYQAPSPPAHSGFHRQYFVYVQEGKVIS-----LLPK 167

DB 85 HFLVVMKGNLSSGTV-----LSDYVSGSPKGTGLHRYVWLVEYQDRPLKCDPEILSNR 140

QY 168 ENKTRGSKMDRFLNRFHLGPEASTQPMQTQ 198

DB 141 SCDHRGKFKVASFRRKKYELRAPVAGTCYQAE 171

RESULT 11

TE26_TOXCA STANDARD; PRT; 262 AA.

AC P54190;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 26 KDA SECRETED ANTIGEN PRECURSOR (TOXOCARA EXCRETORY-SECRETORY

DE ANTIGEN-26) (TES-26).

OS Toxocara canis.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;

OC Toxocaridae; Toxocara.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95355481; PubMed=7629180;

RA Gems D., Ferguson C.J., Robertson B.D., Nieves R., Page A.P.,

RA Blaxter M.L., Maizels R.M.

RT "An abundant, trans-spliced mRNA from Toxocara canis infective larvae

RT encodes a 26-kDa protein with homology to phosphatidylethanolamine-

RT binding proteins."

RL J. Biol. Chem. 270:18517-18522(1995).

CC -1- FUNCTION: BINDS PHOSPHATIDYLETHANOLAMINE.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- DEVELOPMENTAL STAGE: FOUND IN LARVA, BUT NOT IN THE ADULT

CC PARASITE.

CC -1- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING

CC PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch).

CC EMBL; U29761; AAC46843.1; -

CC INTERPRO; IPR001858; -

CC PFAM; PF01161; PBP; 1.

CC PROSITE; PS01220; PBP; 1.

CC Lipid-binding; Antigen; Signal.

CC SIGNAL 1 21

CC FT CHAIN 22 262 26 KDA SECRETED ANTIGEN.

CC SEQUENCE 262 AA; 28048 MW; 426ED0DESCC070AL CRC64;

Query Match 13.8%; Score 169.5; DB 1; Length 262;

Best Local Similarity 27.2%; Pred. No. 4e-09;

Matches 56; Conservative 27; Mismatches 86; Indels 37; Gaps 5;

QY 25 DENSPCAHEALLDEDTLF-----CQGLEVPYFELGNIGCKVVP----- 62

DB 61 DEANCAASINLCQNPTEFPLVRDRCKTCGLCAGCGFSSGIVPLVVTSPSRRVSVTF 120

QY 63 -----DCNNYRKITSWMEPIVKFPGAVDGYILVWYDPDAPSPRAEPQRQFWRHVLVT 116

DB 121 ANNQVNCNGLTTLTAQVANOPTVWE-AQPNDRYTLIMVDPDPSPSAAGQOQORLHWWI 179

QY 117 DIKGADLKKGIQQLSAYQAPSPPAHSGFHRQYFVYVQEGKV-----ISLLPKENKTR 172

DB 180 NIPGNNI-----AGGTTAAFPQSTPAANTGVHRYVELVYRQPAAINSPLLNLNVQDSR 235

QY 173 GSWKMDRFLNRFHLGPEASTQPMQTQ 198

DB 236 PGFGTTAFATQFNLGSPYAGNFYRSQ 261

RESULT 12

OBA5_DROME STANDARD; PRT; 210 AA.

AC P54185;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE PUTATIVE ODORANT-BINDING PROTEIN A5 PRECURSOR (ANTENNAL PROTEIN 5).

GN A5.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CANTON-S; TISSUE=ANTENNA;

RX MEDLINE=94121915; PubMed=7545907;

RA Pikielny C.W., Hasan G., Rouyer F., Rosbash M.;

RT "Members of a family of Drosophila putative odorant-binding proteins

RT are expressed in different subsets of olfactory hairs."

RL Neuron 12:35-49(1994).

CC -1- SUBCELLULAR LOCATION: SECRETED IN THE LUMEN OF THE SENSILLA

CC (POTENTIAL).

CC -1- TISSUE SPECIFICITY: CELLS AT THE BASES OF A FEW SCATTERED SENSILLA

CC ON THE POSTERIOR SURFACE OF THE ANTENNA.

CC -1- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING

CC PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U05243; AAC46472.1; -

CC FLYBASE; FBgn0011294; a5.

CC INTERPRO; IPR001858; -

CC PFAM; PF01161; PBP; 1.

```

KW 3D-structure.
SQ SEQUENCE 181 AA; 20322 MW; 32673415FE29E503 CRC64;

Query Match 12.1%; Score 148; DB 1; Length 181;
Best Local Similarity 29.1%; Pred. No. 2.9e-07;
Matches 44; Conservative 25; Mismatches 62; Indels 20; Gaps 5;

QY 60 VVPDCNN-----YRQKITSWMEPIVKFFGAVDGAFTIYLMVDPDAPSRAEPQR 108
   | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 31 VIYNSNSIKHVYNGHELFPASVTS--TPRVEHVGDMRSFETLMTDPDVPGPSDPYLR 88
   | : | | : | | : | | : | | : | | : | | : | | : | | : | |
QY 109 FWRHLVTDIKG-ADLKKKIQOELSAYQAPSPAINSGHFHYOFFVYLOEGKKVISLLPK 167
   | | | | | | | | | | : | : | : | | : | | : | | : | | : | |
Db 89 EHLHWIWDIPGTITDSSFGK----EVVSYEMPRP--NIGIHRFVFLFKQKKRGQAMLSP 142
   | : | | : | | : | | : | | : | | : | | : | | : | | : | |
QY 168 ENKTRGSKWMDRFLNRFHLGEPASTQFMQT 198
   | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 143 PIVCRDGFNRKFTQENELGLPVAAVFNCQ 173
   | : | | : | | : | | : | | : | | : | | : | | : | | : | |

RESULT 14
PBP_PLAFA
ID PBP_PLAFA STANDARD; PRT; 190 AA.
AC P54189;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PUTATIVE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP MEDLINE=95364848; PubMed=7637711;
RA Trottein F., Cowman A.F.;
RT "The primary structure of a putative phosphatidylethanolamine-binding protein from Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 70:235-239(1995).
CC -! SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U18984; AAC47026.1; -.
DR INTERPRO; IPR001858; -.
DR PFAM; PF01161; PBP; 1.
DR PROSITE; PS01220; PBP; 1.
KW Lipid-binding.
SQ SEQUENCE 190 AA; 21576 MW; 1EA279A1C81CE41 CRC64;

Query Match 11.2%; Score 137.5; DB 1; Length 190;
Best Local Similarity 31.5%; Pred. No. 3.2e-06;
Matches 35; Conservative 20; Mismatches 43; Indels 13; Gaps 3;

QY 86 DGATYIYLMVDPDAPSRAEPQRFRHWLVTDIKGDALKKKIIOG-QELSAYQAPSPPAH 144
   | : | | | | | | | | | | : | | : | | : | | : | | : | |
Db 66 DGVCFLVFWMDPVPRLRDPDKGYIHWVSGIKTKELIKGTQNCVTLIVPGPSIKKG 125
   | : | | | | | | | | | | : | | : | | : | | : | | : | |
QY 145 SGFHRYOFFVYLOEGKKVISLLPENKTR----GSWKMDRFLNRFHLGPEA 191
   | | | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 126 TGLHRISF-----IISLIKEEDKNITGLPHYKGEKYITRVKFNYES 168
   | : | | : | | : | | : | | : | | : | | : | | : | | : | |

RESULT 15
YL79_YEAST
DD YL79 YEAST STANDARD; PRT; 201 AA.

```

Sat Mar 24 13:27:34 2001

```

AC Q06252;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOPHETICAL 22.2 KDA PROTEIN IN TFS1-SAM1 INTERGENIC REGION.
GN YLR179C OR L9470.20.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifkin L., Riles L., Faich A., Trevaskis E., Vignati D., R.;
RA Wilcox L., Wohlman P., Vaudin M., Willson R., Waterston R.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
CC PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U17246; AAB67472.1; -
CC SGD; S0004169; YLR179C.
CC INTERPRO; IPR001858; -
CC PFAM; PF01161; PBP; 1.
CC PROSITE; PS01220; PBP; 1.
CC Hypothetical protein
KW
SQ SEQUENCE 201 AA; 22151 MW; 00157450B02451A2 CRC64;
-----
Query Match 10.9%; Score 133; DB 1; Length 201;
Best Local Similarity 31.0%; Pred. No. 9.2e-06;
Matches 39; Conservative 19; Mismatches 38; Indels 30; Gaps 6;
QY 92 LVMVDPDAPSAEPQRWRHVLVTDI-----KGADLK---KGKIQGQELSAYQAPSPPA 143
DB 75 LLMTDPDAPSRTEHKWSEVCHYIITDIPVEYCGGDIATSGKGVVRNN----YIGPGPPK 130
QY 144 HSGPHRYOFFVYVLOGKVTISLLPKENKTRGSKWMDRFLNRPHLGEP-----EASTQ 194
DB 131 NSGYHRYVFFLCKQ-----PKGADSTTKVENIIS-WGYGTPGAGVDYIKENNLQ 181
QY 195 FMTQNY 200
DB 182 LVGANV 187

```

Search completed: March 24, 2001, 13:24:26
Job time: 172 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:22:28 ; Search time 47.63 Seconds
(without alignments)
162,964 Million cell updates/sec

Title: US-09-215-435-225

Perfect score: 1224

Sequence: 1 MGWTRLVTAALLGLMMVV.....APRRASEPKHKNOAETAAAC 227

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_36.*
- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1224	100.0	227	20 Y35976	Extended human sec
2	1224	100.0	227	21 Y64647	Human phosphatidyl
3	1213	99.1	227	21 Y94263	Human phospholipid
4	652	53.3	121	20 Y11860	Human 5' EST seque
5	235.5	19.2	187	13 R27897	HCNP protein. Rat
6	235.5	19.2	187	15 R49942	Rat hippocampal ch
7	231	18.9	152	21 Y94265	Onchocerca volvulu
8	205	16.7	187	13 R27718	HCNP precursor pro
9	205	16.7	187	15 R49943	Human hippocampal
10	205	16.7	187	16 R64268	Phosphatidylethano
11	164	13.4	175	20 Y49098	Amino acid sequenc
12	148	12.1	181	18 W13944	Antirrhinum centro

13	139	11.4	152	12	R15223	OV-16 antigen. On
14	137.5	11.2	177	18	W13945	Arabidopsis termin
15	89.5	7.3	120	20	Y11503	Human 5' EST seque
16	89.5	7.3	176	19	W72901	Mycobacterium tube
17	89.5	7.3	176	20	Y21918	Amino acid sequenc
18	86.5	7.1	156	20	Y37418	Amino acid sequenc
19	81	6.6	234	12	R13050	CD4-specific CDR-g
20	79.5	6.5	232	16	R80616	Anti-human IL-4 hu
21	77	6.3	110	20	Y35525	Chlamydia pneumoni
22	77	6.3	579	20	Y14130	Human IL-1R18 prot
23	77	6.3	875	19	W37098	Murine endothelial
24	76.5	6.2	236	21	Y56286	HCAT1 clone 25 ant
25	76.5	6.2	321	21	Y92341	Human cancer assoc
26	76.5	6.2	490	21	Y56637	HCAT1 binding huma
27	76	6.2	233	14	R30777	PH52-9.0 humanise
28	75	6.1	585	20	W97757	S-region transfer
29	75	6.1	670	20	Y34517	Porphyromonas ging
30	75	6.1	702	20	Y34390	Porphyromonas ging
31	74.5	6.1	658	20	Y22164	Human TIGIR prote
32	74.5	6.1	686	20	Y14128	Human IL-1R18 prot
33	74	6.0	548	20	Y06927	C. albicans antige
34	73.5	6.0	1192	20	Y23899	Human resenilin bi
35	73.5	6.0	1211	18	W24560	Presenilin-interac
36	73	6.0	458	21	Y81990	Chimeric PCB decom
37	73	6.0	1049	17	W11576	Human phosphatidyl
38	73	6.0	1050	17	W11577	Human phosphatidyl
39	73	6.0	1101	19	W23947	Human phosphoinosi
40	73	6.0	1101	20	W90089	Human G-protein re
41	73	6.0	1101	21	Y78803	Human p120 regulat
42	73	6.0	1711	19	W70506	Osteostesticular pr
43	73	6.0	1711	19	W70507	Mutant osteostestic
44	72.5	5.9	327	21	Y83092	F-box protein FBP-
45	72.5	5.9	503	19	W49909	Signal regulatory

ALIGNMENTS

RESULT 1

Y35976
ID Y35976 standard; Protein; 227 AA.

XX Y35976;

XX AC Y35976;

XX DT 13-SEP-1999 (first entry)

XX DE Extended human secreted protein sequence, SEQ ID NO. 225.

XX KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
cellular differentiation; immune system regulator; anti-inflammatory;
haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
genetic disease.

XX OS Homo sapiens.

XX PN WO9931236-A2.

XX PD 24-JUN-1999.

XX PF 17-DEC-1998; 98WO-IB02122.

XX PR 10-AUG-1998; 98US-0096116.

XX PR 17-DEC-1997; 97US-0069957.

XX PR 09-FEB-1998; 98US-0074121.

XX PR 13-APR-1998; 98US-0081563.

XX PA (GEST) GENSET.

XX PI Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX DR WPI; 1999-385906/32.

XX DR N-PSDB; X97660.

xx PT New isolated human secreted proteins
xx PS Claim 9; Page 255; 516pp; English.
xx

CC This sequence is encoded by an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.

xx SQ Sequence 227 AA;

Query Match 100.0%; Score 1224; DB 20; Length 227;
Best Local Similarity 100.0%; Pred. No. 1.9e-131;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGWMTMLVTAALLGLMMVVTGDEDENSPCAHEALLDEDTLFCQGLEVFPELGNIGCKV 60
DB 1 mgwtmrlvtaaallglmmvvtgdedenspcachealldedtlfcqglevfpeignigckv 60
QY 61 VPCNNYRQKITSMWEPVVKFPGAVDGYATILVMVDPDAPSRAPRFRWHLVLTDIK 120
DB 61 vpcnnyrqkkitswmepivkfpavdgatylvmvdpdapsraepqrfrwhlvtldikg 120
QY 121 ADLKGKIQGQELSAYQAPSPPAHSGFHRFYQFFVYLOEGKVISLLPKENKTRGSKWMDRF 180
DB 121 adlkgkigqgelsayqapsppahsgfhrfyqffvylqegkvisllpkentkrgswkmdrf 180
QY 181 LNRPHLGEPEASTQFMTQNYQDSPTLQAPRERASEPKHKNQAEIAAC 227
DB 181 lnrphlgepeastqfntqnyqdsptlqaprerasepkhknqaeiaac 227

RESULT 2
Y64647
ID Y64647 standard; Protein; 227 AA.

xx AC Y64647;

xx DT 01-FEB-2000 (first entry)

xx DE Human phosphatidylethanolamine-binding protein.

xx KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification.

xx OS Homo sapiens.

xx PN WO953051-A2.

xx PD 21-OCT-1999.

xx PF 09-APR-1999; 99WO-IB00712.

xx PR 09-APR-1998; 98US-0057719.

xx PR 28-APR-1998; 98US-0069047.

xx PA (GEST) GENSET.

xx PI Dumas Milne Edwards J, Duclert A, Giordano J;

xx FT

DR WPI; 2000-038446/03.
DR N-PSDB; 242252.

xx PT Novel secreted protein 5' expressed sequence tag sequences used in
xx diagnostic, forensic, gene therapy, and chromosome mapping procedures
xx
xx Example 21; Page 169-170; 837pp; English.

xx PS
xx
CC 242265 to 243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. Y64651 to Y64658
CC represent the EST-related proteins corresponding to 242265 to 243052.
CC The 5' ESTs can be used for producing secreted human gene products.
CC They can be used to identify and isolate 5' untranslated regions (UTRs)
CC and upstream regulatory regions which control the location, development
CC stage, rate, and quantity of protein synthesis, as well as stability of
CC mRNA. The ESTs are also useful as probes for chromosome mapping, and to
CC obtain full length cDNA clones. The ESTs can also be used in forensic
CC procedures to identify individuals, or in diagnostic procedures to
CC identify individuals having genetic diseases resulting from abnormal
CC gene expression. The products may also be used in gene therapy protocols.
CC The nucleic acids encoding signal peptides can be used for directing
CC extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
CC The proteins encoded by the EST sequences may be useful in treating a
CC variety of human conditions. Secreted proteins have therapeutic value,
CC and the identification of new secreted proteins is valuable. 242249 to
CC 242264 and Y64644 to Y64650 represent sequences used in the
CC exemplification of the present invention.

xx SQ Sequence 227 AA;

Query Match 100.0%; Score 1224; DB 21; Length 227;
Best Local Similarity 100.0%; Pred. No. 1.9e-131;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGWMTMLVTAALLGLMMVVTGDEDENSPCAHEALLDEDTLFCQGLEVFPELGNIGCKV 60
DB 1 mgwtmrlvtaaallglmmvvtgdedenspcachealldedtlfcqglevfpeignigckv 60
QY 61 VPCNNYRQKITSMWEPVVKFPGAVDGYATILVMVDPDAPSRAPRFRWHLVLTDIK 120
DB 61 vpcnnyrqkkitswmepivkfpavdgatylvmvdpdapsraepqrfrwhlvtldikg 120
QY 121 ADLKGKIQGQELSAYQAPSPPAHSGFHRFYQFFVYLOEGKVISLLPKENKTRGSKWMDRF 180
DB 121 adlkgkigqgelsayqapsppahsgfhrfyqffvylqegkvisllpkentkrgswkmdrf 180
QY 181 LNRPHLGEPEASTQFMTQNYQDSPTLQAPRERASEPKHKNQAEIAAC 227
DB 181 lnrphlgepeastqfntqnyqdsptlqaprerasepkhknqaeiaac 227

RESULT 3

Y94263
ID Y94263 standard; protein; 227 AA.

xx AC Y94263;

xx DT 01-AUG-2000 (first entry)

xx DE Human phospholipid binding protein 2, PLBP2.

xx KW Human; phospholipid binding protein; PLBP2; foetal development disorder;
KW reproduction disorder; cell proliferation disorder; immune response;
KW autoimmune disorder; AIDS; infertility; cytostatic; immunosuppressive;
KW gene therapy; hereditary neuropathy;
KW phosphatidylethanolamine binding protein D1; PE-BP D1.

xx OS Homo sapiens.

xx FH key Location/Qualifiers

xx FT Peptide 1..21

FT Modified-site /label= signal_peptide
FT 4
FT /note= "Thr can be phosphorylated by protein kinase C"
FT 21
FT Modified-site /note= "Thr can be phosphorylated by casein kinase II"
FT 67
FT Modified-site /note= "Tyr can be phosphorylated by tyrosine kinase"
FT 73
FT Modified-site /note= "Ser can be phosphorylated by casein kinase II"
FT 101
FT Modified-site /note= "Ser can be phosphorylated by casein kinase II"
FT 169
FT Modified-site /note= "Asn can be glycosylated"
FT 174
FT Modified-site /note= "Ser can be phosphorylated by protein kinase C"
XX US6063767-A.
XX
XX 16-MAY-2000.
XX
XX 09-DEC-1998; 98US-0208718.
XX
XX 28-OCT-1997; 97US-0958820.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Corley NC, Shah P, Lal P, Hillman JL;
XX
XX WPI; 2000-375529/32.
XX N-PSDB; A15382.
XX
XX New purified phospholipid binding proteins 1 and 2 useful for
PT diagnosing, treating or preventing diseases disorders associated with
PT fetal development, reproduction, cell proliferation, and the immune
PT response
XX
XX Disclosure; Fig 2; 37pp; English.
XX
XX The present sequence is the phospholipid binding protein 2 (PLBP2). This
CC protein is expressed in lung, prostate and heart tissues. Also, this
CC protein is expressed in foetal tumour tissues. PLBP2 may be used for the
CC diagnosis, prevention, or treatment of disorders associated with foetal
CC development (e.g. hereditary neuropathies), reproduction (e.g.
CC infertility), cell proliferation (e.g. cancers), and the immune response
CC (AIDS). PLBP2 antibodies may also be developed for potential drug
CC screening or to quantitate PLBP2 gene expression in biopsied tissues. The
CC PLBP2 gene may be administered for gene therapy of disorders associated
CC with PLBP2. PLBP2 has high homology with the phosphatidylethanolamine
CC binding protein DI, PE-BP DI, of Onchocerca volvulus. PE-BP DI is thought
CC to play a role in transport or signal mechanisms between membranes and
CC the cytoplasm.
XX
XX Sequence 227 AA;
XX
Query Match 99.1%; Score 1213; DB 21; Length 227;
Best Local Similarity 99.1%; Pred. No. 3.5e-130;
Matches 225; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MGWTMRLVTAALLGLMMVVTGDENSPCAHEALLDEDTLFCOGLEVFYPELGNICKV 60
DB 1 mgwlmrlvtaallglmmvvtgdedenspcachealldedtlfcgglevfypelgnickv 60
QY 61 VPCDNNYRQKITSWMEPIVKFPGAVDGTATYLLVMVDPDAPSRAEPRQRFWRHVLVTDIKG 120
DB 61 vpcdnnyrqkitswmeplvkfpgavdgatyllyvmvdpdapsraeprqrfwrhvlvtdikg 120
QY 121 ADLKGKIQGELSAYQAPSPPAHSGFHRQYFFVYLOEGKVISILLPKENKTRGSKMDRF 180
DB 121 adlkegiqgelsayqapspahsgfhrqyffvylqegkvissllpkentktrgskmdrf 180
QY 181 LNRPHLGEPEASTQFMONTYQDSFTLQAPRERASEPKHKNQAEIAAC 227
DB 181 lnrphlgepeastqfmontyqdsftlqaprerasepkhknqaeiaac 227

DB 181 lnrphlgepeastqfmontyqdsftlqaprgrasephkhknqaeiaac 227
RESULT 4
Y11860
ID Y11860 standard; Protein; 121 AA.
XX
XX Y11860;
XX
XX 18-JUN-1999 (first entry)
XX
XX Human 5' EST secreted protein SEQ ID No: 460.
DE
XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductuve hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
XX
XX Homo sapiens.
XX
XX WO9906550-A2.
XX
XX 11-FEB-1999.
XX
XX 31-JUL-1998; 98WO-IB01232.
XX
XX 01-AUG-1997; 97US-0905144.
XX
XX (GEST) GENSET.
XX
XX Duclert A, Dumas Milne Edwards J, Lacroix B;
XX WPI; 1999-153780/13.
XX N-PSDB; X40582.
XX
XX New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
XX Claim 34; Page 589; 675pp; English.
XX
XX X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins expressed in prostate, and encode the proteins given in
CC Y11716 to Y11993 respectively. The proteins given represent the signal
CC peptide and an N-terminal fragment of a secreted protein. The nucleic
CC acid sequences can be used for producing secreted human gene products.
CC They can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell proliferation and
CC differentiation activity, haematopoiesis regulating activity, tissue
CC growth regulating activity, reproductuve hormone regulating activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition
CC activity or other activities. The products can be used in forensic, gene
CC therapy and chromosome mapping procedures. The sequences can also be used
CC for obtaining corresponding promoter sequences. The nucleic acids
CC encoding the signal peptides can be used for directing extracellular
CC secretion of a polypeptide or the insertion of a polypeptide into a
CC membrane, or importing a polypeptide into a cell.
XX
XX Sequence 121 AA;
XX
Query Match 53.3%; Score 652; DB 20; Length 121;
Best Local Similarity 99.2%; Pred. No. 1.2e-66;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MGWTMRLVTAALLGLMMVVTGDENSPCAHEALLDEDTLFCOGLEVFYPELGNICKV 60
DB 1 mgwlmrlvtaallglmmvvtgdedenspcachealldedtlfcgglevfypelgnickv 60
QY 61 VPCDNNYRQKITSWMEPIVKFPGAVDGTATYLLVMVDPDAPSRAEPRQRFWRHVLVTDIKG 120

二二二

52 ELGNIG
III .

WMEPIVKFPGAVI

Db 86 hflvnmkngndissgtv----lseyvgsgppkdtglhryvwlvyeqqpcncdepillsnk 141
 QY 168 ENKTRGSGKMDRFLNRFHLGCEPEASTQFMTQYODS 203
 Db 142 sgdnrgkfkvesfrkynhlgapvagtcfgae-wdds 176

RESULT 7.
 ID Y94265 standard; protein; 152 AA.
 AC Y94265;
 KW Phospholipid binding protein; PLBP2; foetal development;
 DT 01-AUG-2000 (first entry)
 DE Onchocerca volvulus phosphatidylethanolamine binding protein D1.
 XX
 KW reproduction; cell proliferation; immune response; cancer; AIDS;
 KW infertility; gene therapy; phosphatidylethanolamine binding protein D1;
 KW PE-BP D1; nematode.
 XX
 OS Onchocerca volvulus.
 XX
 PN US6063767-A.
 XX
 PD 16-MAY-2000.
 XX
 PF 09-DEC-1998; 98US-0208718.
 XX
 PR 28-OCT-1997; 97US-0958820.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Corley NC, Shah P, Lal P, Hillman JL;
 DR WPI; 2000-375529/32.
 XX
 PT New purified phospholipid binding proteins 1 and 2 useful for
 PT diagnosing, treating or preventing diseases disorders associated with
 PT fetal development, reproduction, cell proliferation, and the immune
 PT response -
 XX
 PS Disclosure; Fig 4; 37pp; English.
 XX
 CC The present sequence is the phosphatidylethanolamine binding protein D1,
 CC PE-BP D1 of Onchocerca volvulus. PE-BPs are expressed at a high level in
 CC brain, spleen, lung, liver, skeletal muscle, kidney and testis tissue.
 CC PE-BPs are thought to be essential in normal rat spermatogenesis. PE-BP
 CC D1 is thought to play a role in transport or signal mechanisms between
 CC membranes and the cytoplasm. Nematode PE-BP D1 has high homology with
 CC human phospholipid binding protein 2, PLBP2. PLBP2 is important in foetal
 CC development, reproduction, cell proliferation and the immune response.
 XX
 SQ Sequence 152 AA;

Query Match 18.9%; Score 231; DB 21; Length 152;
 Best Local Similarity 37.0%; Pred. No. 1.6e-18;
 Matches 54; Conservative 18; Mismatches 58; Indels 16; Gaps 5;

QY 55 NIGCKVVP-DCNNYRKITSWMEPIVFKPGAVDGATYILVMVDPDAPSRAPRQRFWRH 113
 Db 20 nigneltqvkngptkv-sw-----daepgaltylmtcdpdknrvfrehwh 70
 QY 114 LVTDIKGADLKKGIQOELSAYQAPSPPAHSGFHRFYQFFVYLOEGKVISLLPKENKTRG 173
 Db 71 liinisgnvssgtv----lsdyigsgppkgtglhryvflvykqpgsitt-tqhggnrp 124
 QY 174 SWKMDRFLNRFHLGCEPEASTQFMTQY 199
 Db 125 nfkvmdfankhhlgnpvagnffqakh 150

RESULT 8
 ID R27718 standard; Protein; 187 AA.
 XX R27718;
 XX
 DT 18-MAR-1993 (first entry)
 DE HCNP precursor protein #2.
 XX
 KW Hippocampal cholinergic neurotrophic peptide; HCNP; hippocampal tissue;
 KW neonatal rat; expression vector; neurodegenerative disorder; dementia;
 KW Alzheimer's disease; Parkinson's disease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 2..12
 FT /label= hHCNP
 FT /note= "This peptide may be truncated to include
 FT amino acids 3-8 or 7-12"
 XX
 PN EP511816-A.
 XX
 PD 04-NOV-1992.
 XX
 PF 27-APR-1992; 92EP-0303800.
 XX
 PR 27-APR-1991; 91JP-0124688.
 XX
 PA (OJIK/) OJIK K.
 PA (SUMU) SUMITOMO PHARM CO LTD.
 PA (YAMA/) YAMAMOTO M.
 XX
 PI Agui H, Fukushima N, Irie T, Kojima S, Nishihara T;
 PI Ojika K, Ono K, Tohdoh N, Tojo S, Ueki Y;
 XX
 DR WPI; 1992-367633/45.
 DR N-PSDB; Q30002.
 XX
 PT Neurotrophic peptide derivs. - used for treating neurological,
 PT degenerative disorders, e.g. Alzheimer's disease or Parkinson's
 PT disease
 XX
 PS Disclosure; Page 42-44; 57pp; English.
 XX
 CC The sequence given is the human hippocampal cholinergic neurotrophic
 CC peptide (HCNP) precursor protein. The active part of this protein
 CC is located in the first eleven amino acids at the N terminus. The
 CC gene encoding this protein was isolated using the rat HCNP precursor
 CC gene sequence as a guide (see also Q30001). From studies of the
 CC structure of the rat HCNP precursor gene and protein the human HCNP
 CC was determined. The rat and human HCNP's have markedly different
 CC structures and composition. HCNP's are useful for the treatment of
 CC neurodegenerative disorders and dementia eg. Alzheimer's and
 CC Parkinson's disease.
 XX
 SQ Sequence 187 AA;

Query Match 16.7%; Score 205; DB 13; Length 187;
 Best Local Similarity 33.1%; Pred. No. 2e-15;
 Matches 50; Conservative 21; Mismatches 62; Indels 18; Gaps 4;

QY 52 ELGNIGCKVVPDCNNYRKITSWMEPIVFKPGAVDGATYILVMVDPDAPSRAPRQRFWR 111
 Db 36 elgkv---ltpqknrptsisw-----dglsgklytlvtdpdparsrkdpykreh 85
 QY 112 HWLVTIDKADLKKGIQOELSAYQAPSPPAHSGFHRFYQFFVYLOEGKVIS----LLPK 167
 Db 86 hflvnmkngndissgtv----lsdyvgsgppkgtglhryvwlvyeqqdrplkcdpillsnr 141

```
QY 168 ENKTRGSKMDRFLNRFHLGEPPEASTQFMTO 198
ID R49943
XX AC R49943;
XX DE 13-OCT-1994 (first entry)
XX KW Human hippocampal cholinergic neurotrophic peptide precursor.
XX DE Human hippocampal cholinergic neurotrophic peptide; human HCNP;
XX KW nerve degeneration; acetylcholine synthesis; neurostimulation.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 2..12
FT /label= human_HCNP
FT Protein 1..187
FT /label= precursor
XX PN W09405788-A.
XX PD 17-MAR-1994.
XX PF 27-AUG-1993; 93WO-JP01214.
XX PR 29-AUG-1992; 92JP-0253734.
XX PR 28-AUG-1992; 92JP-0254170.
XX PA (SUMU) SUMITOMO PHARM CO LTD.
XX PA (YAMA) YAMAMOTO M.
XX PI Agui H, Fukushima N, Irie T, Kamikawa Y, Kojima S;
XX PI Nishihara T, Ojika K, Ono K, Taiji M, Tohdoh N;
XX PI Tojo S, Ueki Y;
XX DR WPI: 1994-101193/12.
XX DR N-PSDB; Q58686.
XX PT Neurotrophic peptide(s), precursors and genes - used to treat
XX PT nervous degeneration, increases acetylcholine synthesis
XX PS Claim 4; Page 164-166; 200pp; Japanese.
XX CC The first 11 amino acids from the N-terminus of the human
XX CC hippocampal cholinergic neurotrophic peptide precursor have
XX CC neurotrophic activity and increase acetylcholine synthesis.
XX SQ Sequence 187 AA;

Query Match 16.7%; Score 205; DB 15; Length 187;
Best Local Similarity 33.1%; Pred. No. 2e-15;
Matches 50; Conservative 21; Mismatches 62; Indels 18; Gaps 4;

QY 52 ELGNIGCKVVPDCNNYRQKITSWMEPIVKFPGAVDGTATYILVMVDPDAPSPRAEPRQRFWR 111
DB 36 elgkv---ltptqvknpstisw-----dglsgklytlvtldpdparkdkpyrewh 85
QY 112 HWLVTDIKGADLKKKGIOGELSAYQAPSPPAHSGFHRVQFFVYLOEGKVIS----LLPK 167
DB 86 hflvnmkgnidssgtv-----lsdyvsgppkgthryvwlvyeqdrplkcdepilsnr 141
QY 168 ENKTRGSKMDRFLNRFHLGEPPEASTQFMTO 198
DB 142 sgdhrgkfkwaskfrkkyelrapvagtcygae 172
```

```
RESULT 10
R64268
ID R64268 standard; Protein; 187 AA.
XX AC R64268;
XX DT 04-JUL-1995 (first entry)
XX DE Phosphatidylethanolamine binding protein.
XX KW Human; phosphatidylethanolamine binding protein; hPEBP; cDNA library;
XX KW glioblastoma; T98G; aplasia; proliferation; glia; neuron;
XX KW hematopoietic cell; immunological activity; neurological activity;
XX KW inflammatory disease; tumour; antibody.
XX OS Homo sapiens.
XX PN EP628631-A.
XX PD 14-DEC-1994.
XX PF 30-MAY-1994; 94EP-0108321.
XX PR 08-JUN-1993; 93JP-0137042.
XX PA (ONOF) ONO PHARM CO LTD.
XX PI Naito T, Nakade S;
XX DR WPI: 1995-015698/03.
XX DR N-PSDB; Q80734.
XX PT New human phosphatidylethanolamine binding protein gene - used to
XX PT develop prods. for diagnosis and treatment of diseases involving
XX PT glia, neurons and other cells.
XX PS Claim 1; Page 13-14; 17pp; English.
XX CC This sequence represents the human phosphatidylethanolamine binding
XX CC protein, hPEBP. The cDNA was isolated from a cDNA library prepared
XX CC using RNA extracted from the human glioblastoma cell line T98G. This
XX CC protein can be used for the prevention of or in the treatment of
XX CC aplasia or abnormal proliferation of glia, neurons or hematopoietic
XX CC cells, depression or enhancement of immunological or neurological
XX CC activity, inflammatory disease, tumours or diseases induced by
XX CC abnormal lipid metabolism. Anti-PEBP antibodies may be used for
XX CC diagnosis of genetic diseases.
XX SQ Sequence 187 AA;

Query Match 16.7%; Score 205; DB 16; Length 187;
Best Local Similarity 33.1%; Pred. No. 2e-15;
Matches 50; Conservative 21; Mismatches 62; Indels 18; Gaps 4;

QY 52 ELGNIGCKVVPDCNNYRQKITSWMEPIVKFPGAVDGTATYILVMVDPDAPSPRAEPRQRFWR 111
DB 36 elgkv---ltptqvknpstisw-----dglsgklytlvtldpdparkdkpyrewh 85
QY 112 HWLVTDIKGADLKKKGIOGELSAYQAPSPPAHSGFHRVQFFVYLOEGKVIS----LLPK 167
DB 86 hflvnmkgnidssgtv-----lsdyvsgppkgthryvwlvyeqdrplkcdepilsnr 141
QY 168 ENKTRGSKMDRFLNRFHLGEPPEASTQFMTO 198
DB 142 sgdhrgkfkwaskfrkkyelrapvagtcygae 172

RESULT 11
Y49098
ID Y49098 standard; Protein; 175 AA.
XX
```


FT Modified-site 56..58 /note= "N-linked glycosylation site"
FT Modified-site 61..63 /note= "N-linked glycosylation site"
FT Modified-site 119..121 /note= "N-linked glycosylation site"
FT Modified-site 124..126 /note= "N-linked glycosylation site"

XX US7644372-A.
XX
XX
XX 05-NOV-1991.
XX
XX 23-JAN-1991; 91US-0644372.
XX
XX 23-JAN-1991; 91US-0644372.
XX (USSH) NAT INST OF HEALTH.
XX
XX Lazzeri MESL, Nutman TB, Weiss N;
XX
XX WPI; 1991-361687/49.
XX N-PSDB; Q14832.
XX
XX Onchocerca volvulus antigen (OV-16) and DNA - useful in early and
PT specific diagnosis of onchocerciasis
XX
XX Disclosure: Fig 4; 48pp; English.
XX
XX The amino acid sequence is that of the Onchocerca volvulus antigen
CC (OV-16) which may be used in the diagnosis of onchocerciasis in an
CC animal. Early and specific diagnosis of new or re-infections with O.
CC volvulus in vector reinvasion areas, as well as the detection of light
CC infections in areas where control is being attempted by widespread use
CC of ivermectin, is possible. This allows for monitoring, evaluating and
CC consolidating onchocerciasis control by both the vector control and
CC chemotherapeutic strategies.
XX
XX Sequence 152 AA;

Query Match 11.4%; Score 139; DB 12; Length 152;
Best Local Similarity 38.5%; Pred. No. 4.9e-08;
Matches 35; Conservative 9; Mismatches 33; Indels 14; Gaps 4;
QY 55 NTGCKVVP-DCNNYRKITSWEPVVKFGAVDGYIILVMVDPDAPSRAEPRQRFRWHW 113
Db 65 nlgneltptqvknpktkv-sw-----daegpalytlvmtddapsrknvpvfwfwhhw 115
QY 114 LVTDIKADLKKGIQOELSAYOAPSPPAH 144
Db 116 llinisgqvnsqgtv---lsdywirstkrh 142

RESULT 14
W13945
ID W13945 standard; Protein; 177 AA.
XX
XX W13945;
XX
XX 24-JUN-1997 (first entry)
XX Arabidopsis terminal flower 1 protein.
XX Arabidopsis terminal flower 1 protein.
XX Terminal flower 1; tfl1 gene; centroradialis protein; cen gene;
KW flowering; transgenic plant.
XX
XX Arabidopsis thaliana var. Columbia.
XX
XX W09710339-A1.
XX
XX 20-MAR-1997.
XX

PF 13-SEP-1996; 96WO-GB02276.
XX
PR 13-SEP-1995; 95GB-0018731.
XX
XX (INNE-) INNES CENT JOHN.
XX
XX Bradley DJ, Carpenter R, Coen ES;
PI
XX WPI; 1997-202235/18.
DR N-PSDB; T60140;
DR N-PSDB; T60142.

XX Antirrhinum majus centroradialis gene and Arabidopsis homologue,
PT tfl1 - control switching of apical meristem to floral fate, useful
PT in plant hybridisation and in control of growing season
XX
XX Claim 12; Fig 6a; 83pp; English.

XX The terminal flower 1 (TFL1) protein (W13945) of Antirrhinum majus
CC plays a role in preventing the apical meristem from switching to
CC a floral fate and in inhibition of flowering. Its sequence can be
CC deduced from an isolated cDNA clone (T60140) and genomic sequence
CC (T60142). Mutants and variant TFL1 polypeptides, and homologous
CC CSN protein (see also W13944) of Antirrhinum, can be expressed in
CC transgenic plants in order to influence flowering characteristics,
CC esp. the switching of apical meristem to a floral fate and the
CC timing of flowering. Antibodies raised against the polypeptides
CC are useful in the identification and isolation of homologous
CC proteins.
XX
XX Sequence 177 AA;

Query Match 11.2%; Score 137.5; DB 18; Length 177;
Best Local Similarity 30.2%; Pred. No. 8.9e-08;
Matches 38; Conservative 25; Mismatches 50; Indels 13; Gaps 5;
QY 76 EPIVKFPGAVDGYIILVMVDPDAPSRAEPRQRFRWHWLVTDIKG-ADLKKGIQOELS 134
Db 54 kprveihggdlrsfftlvmidpdpvpsdpflkehlhwtvntnpgttatfgk----evv 109
QY 135 AYQAPSPPAHSGFHRHYOFFVYLOEGK--VISLLPKENKTGSKWMDRFLNRFHLCEPEAS 192
Db 110 syelprpsi--glhrfvfvrqkrvifpnip-----srdfntrkfaveydlglpva 163
QY 193 TQFMTQ 198
Db 164 vffnaq 169

RESULT 15
Y11503
ID Y11503 standard; Protein; 120 AA.

XX Y11503;
AC
XX
XX 21-JUN-1999 (first entry)
XX Human 5' EST secreted protein SEQ ID No 325.
XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
XX
XX Homo sapiens.
XX
XX W09906551-A2.
XX
XX 11-FEB-1999.
XX

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:25:43 ; Search time 73.46 Seconds
(without alignments)
362.186 Million cell updates/sec

Title: US-09-215-435-225
Perfect score: 1224
Sequence: 1 MGWTRLVTAALLGLMMV.....APRERASEPKHKNAETAAC 227

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_15:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_prodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	229	18.7	178	5 Q9VK60	Q9VK60 drosophila
2	225	18.4	187	11 Q9J58	Q9J58 mus musculus
3	222.5	18.2	187	5 Q9VI08	Q9VI08 drosophila
4	216.5	17.7	202	5 Q9VI09	Q9VI09 drosophila
5	206.5	16.9	173	10 Q9XFK7	Q9XFK7 arabisidopsis
6	196	16.0	171	5 Q9NKY4	Q9NKY4 dirofilaria
7	186.5	15.2	179	5 Q9VD02	Q9VD02 drosophila
8	183	15.0	172	5 Q9XW37	Q9XW37 caenorhabdi
9	182	14.9	175	10 Q9S7R5	Q9S7R5 arabisidopsis
10	182	14.9	210	5 Q9VP29	Q9VP29 drosophila
11	179	14.6	177	10 Q9SDY9	Q9SDY9 citrus unsh
12	175.5	14.3	176	5 Q9VD01	Q9VD01 drosophila
13	164	13.4	174	10 Q9XH44	Q9XH44 nicotiana t
14	164	13.4	175	10 Q04467	Q04467 arabisidopsis
15	159	13.0	182	5 Q9Y1K8	Q9Y1K8 octopus dof
16	154.5	12.6	173	10 Q9XGS4	Q9XGS4 oryza sativ
17	152	12.4	173	10 Q9XGS5	Q9XGS5 oryza sativ
18	148.5	12.1	175	10 Q82088	Q82088 lycopersico
19	147.5	12.1	175	10 Q9XH43	Q9XH43 nicotiana t

20	147	12.0	175	10 Q9ZNV5	Q9ZNV5 arabisidopsis
21	144.5	11.8	175	10 Q9XH42	Q9XH42 nicotiana t
22	138	11.3	118	10 Q9XH41	Q9XH41 nicotiana t
23	137.5	11.2	177	10 P93003	P93003 arabisidopsis
24	134.5	11.0	177	10 Q9SAY4	Q9SAY4 brassica ol
25	134.5	11.0	178	10 Q82152	Q82152 brassica na
26	134.5	11.0	178	10 Q82153	Q82153 brassica ra
27	132.5	10.8	178	10 Q82153	Q82153 brassica ca
28	132.5	10.8	178	10 Q9XGD4	Q9XGD4 brassica ca
29	122.5	10.0	178	10 Q9SAY6	Q9SAY6 brassica na
30	115.5	9.4	177	10 Q82154	Q82154 brassica ol
31	113	9.2	215	3 Q9P6X9	Q9P6X9 neurospora
32	103	8.4	409	5 Q9V748	Q9V748 drosophila
33	101	8.3	122	10 Q9SXZ2	Q9SXZ2 arabisidopsis
34	96.5	7.9	150	2 Q9JQ64	Q9JQ64 chlamydia p
35	93.5	7.6	157	1 Q9UXF4	Q9UXF4 sulfolobus
36	93.5	7.6	186	1 Q9UJZ3	Q9UJZ3 pyrococcus
37	93	7.4	180	2 Q9KT30	Q9KT30 vibrio chol
38	90.5	7.4	209	2 Q9KJE6	Q9KJE6 thauera aro
39	89.5	7.3	1945	5 Q96681	Q96681 drosophila
40	89.5	7.3	1945	5 Q9V491	Q9V491 drosophila
41	88	7.2	174	4 Q9P0B9	Q9P0B9 homo sapien
42	87	7.1	150	2 Q9PLJ0	Q9PLJ0 chlamydia m
43	85	6.9	449	4 Q9UJH6	Q9UJH6 homo sapien
44	85	6.9	595	4 Q9Y5X1	Q9Y5X1 homo sapien
45	85	6.9	2253	13 P70012	P70012 xenopus lae

ALIGNMENTS

RESULT	1				
Q9VK60					
ID	Q9VK60	PRELIMINARY;	PRT;	178 AA.	
AC	Q9VK60;				
DT	01-MAY-2000 (Tremblrel. 13, Created)				
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)				
DT	01-OCT-2000 (Tremblrel. 15, Last annotation update)				
DE	CG6180 PROTEIN.				
GN	CG6180.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BERKELEY;				
RX	MEDLINE=20196006; PubMed=10731132;				
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,				
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,				
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,				
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,				
RA	Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,				

Dd 36 ELGKV---LTPQVNNRPSSISW-----DGLDPGKLYTLVTDDPAFSRDKPFRWH 85

Qy 112 HVLAVTDIKADLKKGIOQELSAVOAPSPPAHSGFHYOFFYVLOEGKYVIS----LLPK 167
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 86 HFUVNKKGNDSSTGV-----LSDYVGSPPSGTGLHRYWLVIYEQEQLSCDEPLSNK 141
|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 168 ENKTRGSWMKDRLNFRLFHGPEASTOPMTQ 198
||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 142 SGDNRCRKFKVETFRKKYNLGPAVAGTCYQAE 172
||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 3
Q9VII08 PRELIMINARY; PRT; 187 AA.
ID OSVI08
AC QSVI08;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
CD CG10298 PROTEIN.
DE NCBI_TaxID=7227;
GN CG10298.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Permygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephyroridae; Drosofilidae; Drosofila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantadas P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
RA George R.A., Lewis S.E., Richards S.S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blasez R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Held G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottilier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke K., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I.J., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertilla W.S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hariss N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lakso P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J.J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K.K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassafman D.A., Weinstock G.M., Weissbach J.J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003675; AF54137.1; -
DR HSSP; p13696; 1B7A.
DR FLXBASE; FBgn0037432; CG10298.
DR INTRAPRO; IPRO01858; -
DR PFAM; PF01161; PBP; 1.

DR PROSITE; PS01220; PBP; 1.
SQ SEQUENCE 187 AA; 20872 MW; A3C865A04EEACC01 CRC64;

Query Match 18.2%; Score 222.5; DB 5; Length 187;
Best Local Similarity 38.6%; Pred. No. 1.5e-14;
Matches 49; Conservative 21; Mismatches 48; Indels 9; Gaps 3;

QY 76 EPIVKKPGVAGDGYILVMVDPPAPRAEPRQRFWRHLVTDIKGADLKKKGIOGELS 135
DB 50 QPKVKW-DADPNFYTLTLLDPPAPRKKEPKFKEWHHLVNPQGVGVV-----LTE 104

QY 136 YQAPSPPAHSGFHYQFFVYLOGKVISLIPKKNKT-----RGSWKMDRFLNRHFLGEPE 191
DB 105 YVGAGPQGGLHRYVFLFKQPKLTKNEPKIPKTSKGRANFSTSKFNKSKYKLGDP 164

QY 192 STQFMQ 198
DB 165 GNFFQAQ 171

RESULT 4
Q9VI09 PRELIMINARY; PRT; 202 AA.
AC Q9VI09;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE CGI7919 PROTEIN.
GN CGI7919.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananthites P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacleeb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaverl J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003675; AAF54136.1; -.
DR HSSP; P30086; 1BEH.
DR FLYBASE; FBgn0037433; CGI7919.
DR INTERPRO; IPR001858; -.
DR PFAM; PF01161; PBP; 1.
DR PROSITE; PS01220; PBP; 1.
SQ SEQUENCE 202 AA; 22647 MW; F0333A13597AC596 CRC64;

Query Match 17.7%; Score 216.5; DB 5; Length 202;
Best Local Similarity 28.0%; Pred. No. 6.8e-14;
Matches 59; Conservative 31; Mismatches 72; Indels 49; Gaps 6;

QY 13 LLGLMMVVTGDEDENSPCAHEALDEDTLFCQGLEVFYFELGNIGCKVVPDCNNYRK 72
DB 8 LVGCLLAVQAGSVVEVFRSHQVVD-----VIPEPPNQLKVT 45

QY 73 SWM-----EPIVKFGVAGDGYILVMVDPPAPRAEPRQRFWRHLV 115
DB 46 YSNLVAKDGVLTPTQVKDQPVVW-DAQPGFEYTLTMDPPAPRAEPRQRFWRHLV 104

QY 116 TDIKGADLKKKGIOGELSAYQAPSPPAHSGFHYQFFVYLOGKV---ISLLPKENKT 171
DB 105 ANTAGNDL-----ASGEPIAEYIGSGPGQGLHRYVFLYKQSKLEFDEERVKSRK 160

QY 172 RGSWKMDRFLNRHFLGEPEASTQFMQYQD 202
DB 161 RPKFSAKFAINHELGNIPAGTFYQAQ-YDD 190

RESULT 5
Q9XFK7 PRELIMINARY; PRT; 173 AA.
AC Q9XFK7;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE EI2A11 PROTEIN (T10F20.11 PROTEIN) (T10022.8).
GN EI2A11 OR T10F20.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC KARDALLSKY I., Weigel D.;
RA "A new Arabidopsis member of the TFL1/FT gene family.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Sakano H., Liu S.X., Etgu P., Lee J.M., Lenz C., Pham P., Toriumi M.,
RA Yu G., Chan A., Chung M., Goldsmith A., Liu A., Smith A., Vaysberg M.,
RA Altatfi H., Brooks S., Buehler E., Chao Q., Conn L., Conway A.B.,
RA Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B., Miranda M.,
RA Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
RA Federspiel N.A., theologis A.;
RT "The sequence of BAC T10F20 from Arabidopsis thaliana chromosome 1.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altatfi H., Bel Q., Chin C., Chiu J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;

AC	Q9VD02;	Created
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)

Qy	76	EPVKEFGVAGATYI--LVMDPDPAPSRAEPRORERHWLVTDIKGADKKGIQOQEL	13
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	41	QPIVMSG-LEGSKNLTLLWDPDPATRODPKYREILHWSVVNIPGSN--ENPSGGHSL	97
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Qy	134	SAYQAPSPAHSGFHYQYFVYLQEGKVISLLPKENKTRG---SWKMDFLNRFHLGEPE	150
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	98	ADYVGSGPKDTGLHRYIFLLYRQENKRIETPTISNTTRTGRUNFARFAAKHGLCEPI	157
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Qy	191	ASTQFMNTQNYQD	202
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	158	AANYQAOQ-YDD	168
		: : : : : : : : : : : : : : : : : : : : : : : : :	

RESULT	9	
Q9S7R5	PRELIMINARY;	PRT; 175 AA.
ID	Q9S7R5	
AC	Q9S7R5;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)	
DE	TP11 LIKE PROTEIN.	
GN	F9F13.20 OR TSF OR AT4G20370.	
OS	Arabidopsis thaliana (Mouse-ear cress).	
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;	
OC	Brassicales; Brassicaceae; Arabidopsi.	
OX	NCBI_Taxid=3702;	
RN	[1]	
RP	SEQUENCE FROM N.A.	

	Query Match	14.9%	Score 182;	DB 10;	Length 175;
	Best Local Similarity	36.2%;	Pred. No. 1.6e-10;		
	Matches	46;	Conservative 15;	Mismatches 50;	Indels 16; Gaps 4;
Qy	76	EPIVKFGAVDGTATYILVMVDPDAPSRAEPQRQFRWHLVTDIAGADLUKKGIOGQLSA	135		
	:	:			:
Db	51	KPIVEIGDDDFRNEYTLVMVDPDVPSPSNHQREYLHWLVTIDIPAT---	TGNAFGNEVC	107	
	:	:			:
Qy	136	YOAPSPAHSGFHRVQFFVYLQGEKVISLLPKENKTRGSWKMD---	RFLNRFHLGPEEA	191	
	:	:		:	
Db	108	YESRPP--SGIHRIVLVFLQRLGRQTVYAP-----	GWRQQFNTRFAFIYNLGLPVA	158	
	:	:		:	
Qy	192	STQFMQT	198		
	:				
Db	159	ASVENCO	165		
	:				

RESULT	10
Q9VPZ9	
ID	Q9VPZ9
AC	PRELIMINARY;
DC	PRT; 210 AA.
DT	Q9VPZ9;
DT	01-MAY-2000 (TRENBLrel. 13, Created)
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT	01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE	A5 GENE PRODUCT.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Prervota; Neoptera; Endopterygota; Diptera; Brachycera; Muscivora;

RESULT	11
Q9SXY9	
IN	Q9SXY9

Db 65 YTLVMDPDPSPNSPHLEHLVLTDPAT---TGTFEGNEIVCYENPSPTA--GIHR 119
QY 150 YQFFVYLQGGKVISLLPKENKTRGSKMD-----RFLNRFHLGPEASTQFMTO 198
Db 120 VVFLPRQLGRQTVXAP-----GWRQNFREFAEIYNLGLPVAAVFYNCQ 165

RESULT 15
ID Q9YIK8 PRELIMINARY; PRT; 182 AA.
AC Q9YIK8;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE O-CRYSTALLIN.
OS Octopus dofleini (Giant octopus).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Octopoda;
OC Incirrata; Octopodidae; Octopus.
OX NCBI_TaxID=6644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99280325; PubMed=10350626;
RA Zinovieva R.D., Piatigorsky J., Tomarev S.I.;
RT "O-Crystallin, arginine kinase and ferritin from the octopus lens."
RL Biochim. Biophys. Acta 1431:512-517(1999).
DR EMBL; AF117272; RAD29640.1; -;
DR HSSP; Q41261; IQ00.
DR INTERPRO; IPR001858; -;
DR PFAM; PF01161; PBP; 1.
SQ SEQUENCE 182 AA; 21241 MW; D75DB3CC087B22AD CRC64;

Query Match 13.0%; Score 159; DB 5; Length 182;
Best Local Similarity 30.7%; Pred. No. 3.3e-08;
Matches 43; Conservative 25; Mismatches 48; Indels 24; Gaps 6;
QY 76 EPIVKFPGAVDGYTLVMDPDPSPNSPHLEHLVLTDPAT---TGTFEGNEIVCYENPSPTA--GIHR 119
Db 45 QPQIKFE-AETNYYIILNDADFPSPRSQKLNFEQHLVWVNIIPGSDISRGDV----LTD 99
QY 136 YQAPSPPAHSGFHYQFFVYLQ-----EGKVISLLPKENKT---RGSWKMDRFLNR 183
Db 100 YIGPLPNKGTGYHRYVLMFLFKQSKGRMEFRGEKKI-----NNRTSEGRKSYNNMEFARK 153
QY 184 FHLGPEASTQFMTOYQDS 203
Db 154 HFLVERPYGNFFQSE-WDDS 172

Search completed: March 24, 2001, 13:25:45
Job time: 251 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:23:16 ; Search time 44.26 Seconds
(without alignments)
348.248 Million cell updates/sec

Title: US-09-215-435-225

Perfect score: 1224

Sequence: 1 MGWTMLVTAALLGLMMVV.....APRRASPKHKNAQIAAC 227

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_66:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	235.5	19.2	187	2 A36126	23K phosphatidylylethanolamine-binding protein - rat
2	234.5	19.2	221	2 T31721	hypothetical prote
3	223	18.2	134	2 PC4214	phosphatidylylethano
4	222	18.1	152	2 PC4216	phosphatidylylethano
5	218.5	17.9	186	2 S00056	basic cytosolic pr
6	218.5	17.9	186	2 S18358	23K protein - rat
7	210	17.2	114	2 PC4215	phosphatidylylethano
8	206	16.8	187	2 S46485	phosphatidylylethano
9	205	16.7	187	2 T53745	phosphatidylylethano
10	183	15.0	172	2 T27310	hypothetical prote
11	182	14.9	175	2 T10579	TfL1 protein homol
12	169.5	13.8	262	2 A57391	TcSL-2 protein pre
13	148.5	12.1	175	2 T07877	self-pruning prote
14	148	12.1	181	2 S68507	centroradialis pro
15	139	11.4	152	2 A44947	16k antigen precu
16	137.5	11.2	177	2 T48411	terminal flower1 (
17	133	10.9	201	2 S51424	hypothetical prote
18	122.5	10.0	219	2 S18843	TfS1 protein - yea
19	113.5	9.3	197	2 C70519	hypothetical prote
20	113	9.2	151	2 E69134	conserved hypotet
21	113	9.2	215	2 T48789	TfS1 related prote
22	106	8.7	171	2 B70408	conserved hypotet
23	105	8.6	198	2 B71072	hypothetical prote
24	96.5	7.9	150	2 C72024	conserved hypotet
25	95	7.8	183	2 G64786	ybCL protein - Esc
26	95	7.8	201	2 D70519	probable lppc prot
27	93.5	7.6	186	2 C75095	hypothetical prote
28	93	7.6	180	2 C82243	conserved hypotet
29	92	7.5	179	2 T36855	conserved hypotet

30	91.5	7.5	551	2 D69282	glutamyl-tRNA synt
31	89.5	7.3	176	2 H70577	hypothetical prote
32	89.5	7.3	1945	2 T13937	plexin A - fruit f
33	87	7.1	150	2 F81740	conserved hypotet
34	86.5	7.1	150	2 G71477	hypothetical prote
35	85	6.9	2253	2 T30336	nuclear/mitotic ap
36	80.5	6.6	287	2 A69462	conserved hypotet
37	80	6.5	396	2 S76241	probable dioxigena
38	79	6.5	265	2 S00033	histone H1.b - gar
39	78.5	6.4	182	2 T48152	hypothetical prote
40	78.5	6.4	420	2 T14166	fxud protein - Myc
41	78	6.4	390	2 S71197	hypothetical prote
42	78	6.4	406	2 S71242	3'-phosphoadenosin
43	77.5	6.3	455	2 T33572	hypothetical prote
44	77.5	6.3	490	2 A35546	muscarinic acetylch
45	77	6.3	1622	2 JE0378	DNA (cytosine-5') -

ALIGNMENTS

RESULT 1

A36126

23K phosphatidylylethanolamine-binding protein - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 20-Jun-2000

C:Accession: A36126; S46484; S37556; I84473; S37555

R:Grandy, D.K.; Hanneman, E.; Bunzow, J.; Shih, M.; Machida, C.A.; Bidlack, J.M.; Civ

Mol. Endocrinol. 4, 1370-1376, 1990

A:Title: Purification, cloning, and tissue distribution of a 23-kDa rat protein isola

A:Reference number: A36126; MUID:91042640

A:Accession: A36126

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-187 <GRA>

R:Perry, A.C.F.; Hall, L.; Bell, A.E.; Jones, R.

Biochem. J. 301, 235-242, 1994

A:Title: Sequence analysis of a mammalian phospholipid-binding protein from testis an

A:Reference number: S46484; MUID:94311839

A:Accession: S46484

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-187 <PER>

A:Cross-references: EMBL:X71873; NID:g510338; PIDN:CAA50708.1; PID:g510339

R:Tohdo, N.; Agui, H.; Ojika, K.

submitted to the EMBL Data Library, September 1993

A:Description: Sequences of rat HCNP precursor protein promoter region.

A:Reference number: S37556

A:Accession: S37556

A:Molecule type: DNA

A:Residues: 1-45 <TOD>

A:Cross-references: EMBL:X75254

R:Tohdo, N.; Tojo, S.; Agui, H.; Ojika, K.

Brain Res. Mol. Brain Res. 30, 381-384, 1995

A:Title: Sequence homology of rat and human HCNP precursor proteins, bovine phosphati

A:Reference number: I38109; MUID:95364631

A:Accession: I84473

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-187 <RES>

A:Cross-references: EMBL:X75253; NID:g406291; PIDN:CAA53032.1; PID:g406292

C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match 19.2% Score 235.5; DB 2; Length 187;
Best Local Similarity 36.5% Pred. No. 3.4e-16;
Matches 57; Conservative 20; Mismatches 60; Indels 19; Gaps 5;

Oy 52 ELGNIGCKVVPDCNNYRQKITSNMEPIVKFPGAVDGATYILVMVDDPAPSRAEPRQFWR 111

Db 36 ELGKV---LTPQVMNRPSISW-----DGLDPGKLYTLVLTDPOAPSRKDKFREW 85

Oy 112 HNLVTDIKGADLKKGKTQGSLSAYQAPSPAHSGFHRYQFFVYLQEGKVLS----LLPK 167

```
Db 86 HFLVNMKNDLSSGTV-----LSEYVSGPPKDTGLHRYVWLVIYEQEQLNCDPEILSNK 141
QY 168 ENKTRGSKMDRFLNRFHLGEPPEASTQFMQYQDS 203
Db 142 SGDNRGKFKVESFRKKYHLGAPVAGTCFAE-WDDS 176

RESULT 2
T31721
hypothetical protein F40A3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jul-2000
C:Accession: T31721
R:Erttmann, K.D.; Bradshaw, H.; Keppler, D.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid F40A3.
A:Reference number: Z21074
A:Accession: T31721
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-221 <GET>
A:Cross-references: EMBL:AF016423; PIDN:AAB65322.1; GSPDB:GN00023; CESP:F40A3.3
A:Experimental source: strain Bristol N2; clone F40A3
C:Genetics:
A:Gene: CESP:F40A3.3
A:Map position: 5
A:Introns: 32/2; 70/3
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match 19.2%; Score 234.5; DB 2; Length 221;
Best Local Similarity 33.1%; Pred. No. 5.3e-16;
Matches 58; Conservative 16; Mismatches 60; Indels 41; Gaps 6;
QY 59 KVVDP--CNNYRQKITSWMEPIVKFPGAVD-----GATYIL 92
Db 46 EVIPDVLASPPSKVVS-----VKFNSGEANLGNLTPTQVKDTPVKWDAEPGALYTL 100
QY 93 VMVDPDAPSAEPQRQFWRHLVTDIKGADLKKGIQGOELSAQAPSPAHSGFHRHYOF 152
Db 101 IKYDPPDAPSRKEPTYRWHHLVVNIPGNDIAK----GDTLSEYIAGPPPKTGLHRYVY 156
QY 153 FVYLQEGKVIS-----LLPKENKTRGSKMDRFLNRFHLGEPPEASTQFMQYQD 202
Db 157 LIYKQSGRIEDAEHRLTNTSGDKRGWKAADFVAKHKLGAPEVGNLFAE-YDD 210

RESULT 3
PC4214
phosphatidylethanolamine binding protein Ovd3 - nematode (Onchocerca volvulus) (fragment)
C:Species: Onchocerca volvulus
C:Date: 17-Dec-1996 #sequence_revision 21-Jan-1997 #text_change 20-Jun-2000
C:Accession: PC4214
R:Erttmann, K.D.; Gallin, M.Y.
Gene 174, 203-207, 1996
A:Title: Onchocerca volvulus: Identification of cDNAs encoding a putative phosphatidyl-
A:Reference number: PC4214; MUID:97045813
A:Accession: PC4214
A:Molecule type: mRNA
A:Residues: 1-134 <ERT>
A:Cross-references: EMBL:X87989; NID:g1143530; PID:g1143531
C:Genetics:
A:Introns: 34/3; 52/2; 81/1; 129/3
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match 18.2%; Score 223; DB 2; Length 134;
Best Local Similarity 36.3%; Pred. No. 4e-15;
Matches 53; Conservative 18; Mismatches 59; Indels 16; Gaps 5;
QY 55 NIGCKVVP--DCNNYRQKITSWMEPIVKFPGAVDGTATYLVWVDPDAPSAEPQRQFWRHW 113
```

```
Db 2 NLGNELTPTQVKNQTKV-SW-----DAEPGALYTLVWTDPAEPSRKPNPVRWHHW 52
QY 114 LVTDIKGADLKKGIQGOELSAQAPSPAHSGFHRHYOFFVYVYLGEGKVISLLPKENKTRG 173
Db 53 LIINISGQNVSSGTV-----LSDYIGSGOPKGTGLHRYVFLVYKQPGSITD--TQGGNRP 106
QY 174 SWKMDRFLNRFHLGEPPEASTQFMQYQD 199
Db 107 NFKVMDFAFKHHLGNPVGAGNFFQAKH 132

RESULT 4
PC4216
phosphatidylethanolamine binding protein Ovd1 - nematode (Onchocerca volvulus) (fragm
C:Species: Onchocerca volvulus
C:Date: 17-Dec-1996 #sequence_revision 21-Jan-1997 #text_change 20-Jun-2000
C:Accession: PC4216
R:Erttmann, K.D.; Gallin, M.Y.
Gene 174, 203-207, 1996
A:Title: Onchocerca volvulus: Identification of cDNAs encoding a putative phosphatidyl
A:Reference number: PC4214; MUID:97045813
A:Accession: PC4216
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-152 <ERT>
A:Cross-references: EMBL:X87991
C:Comment: This protein is involved in odorant binding and is located in the chemosen
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match 18.1%; Score 222; DB 2; Length 152;
Best Local Similarity 36.3%; Pred. No. 5.9e-15;
Matches 53; Conservative 18; Mismatches 59; Indels 16; Gaps 5;
QY 55 NIGCKVVP--DCNNYRQKITSWMEPIVKFPGAVDGTATYLVWVDPDAPSAEPQRQFWRHW 113
Db 20 NLGNELTPTQVKNQTKV-SW-----DAEPGALYTLVWTDPAEPSRKPNPVRWHHW 70
QY 114 LVTDIKGADLKKGIQGOELSAQAPSPAHSGFHRHYOFFVYVYLGEGKVISLLPKENKTRG 173
Db 71 LIINISGQNVSSGTV-----LSDYIGSGOPKGTGLHRYVFLVYKQPGSITD--TQGGNRP 124
QY 174 SWKMDRFLNRFHLGEPPEASTQFMQYQD 199
Db 125 NFKVMDFAFKHHLGNPVGAGNFFQAKH 150

RESULT 5
S00056
Basic cytosolic protein, 21K - bovine
N:Alternate names: phosphatidylethanolamine-binding protein
C:Species: Bos primigenius taurus (cattle)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 20-Jun-2000
C:Accession: S00056; S51191
R:Schoentgen, F.; Saccoccio, F.; Jolles, J.; Bernier, I.; Jolles, P.
Eur. J. Biochem. 166, 333-338, 1987
A:Title: Complete amino acid sequence of a basic 21-kDa protein from bovine brain cyt
A:Reference number: S00056; MUID:87275914
A:Accession: S00056
A:Molecule type: protein
A:Residues: 1-186 <SCH>
R:Buquoy, S.; Jolles, P.; Schoentgen, F.
Eur. J. Biochem. 225, 1203-1210, 1994
A:Title: Relationships between molecular interactions (nucleotides, lipids and protei
A:Reference number: S51191; MUID:95045550
A:Contents: annotation
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5
C:Keywords: brain; cytosol; lipid binding; nucleotide binding; phospholipid
F:112-125/Region: nucleotide binding
F:168-182/Region: lipid binding
```


QY 173 GSWKMDRFLNRHFLGEPEASTQFMQ 198
: | :|:| | | : :|
Dh 236 PGEGTFAFATQENLGSFYAGNFYRSQ 261

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:23:53 ; Search time 34.59 seconds
(without alignments)
117.845 Million cell updates/sec

Title: US-09-215-435-225
Perfect score: 1224
Sequence: 1 MGWTRLVTAALLIGLMWV.....APRRASEPKHKNOAETAA 227

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2.6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2.6/ptodata/2/1aa/6_COMB.pep:*
4: /cgn2.6/ptodata/2/1aa/PctUS_COMB.pep:*
5: /cgn2.6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1213	99.1	227	3	US-09-208-718-3
2	231	18.9	152	3	US-09-208-718-6
3	210.5	17.2	187	1	US-08-403-378B-4
4	205	16.7	187	1	US-08-403-378B-15
5	139	11.4	152	1	US-07-644-372-2
6	79.5	6.5	232	1	US-08-704-744-80
7	77	6.3	875	1	US-08-785-241-5
8	76	6.2	233	2	US-07-934-373C-25
9	76	6.2	233	3	US-08-437-642B-25
10	76	6.2	233	4	PCT-US93-07832-25
11	73.5	6.0	212	2	US-08-916-770-3
12	73.5	6.0	212	2	US-09-102-593-3
13	73	6.0	1049	2	US-08-817-090B-2
14	73	6.0	1050	2	US-08-817-090B-4
15	73	6.0	1101	2	US-08-916-917-14
16	73	6.0	1101	3	US-09-225-170-14
17	73	6.0	1711	2	US-08-342-930-2
18	71.5	5.8	421	2	US-08-576-626A-53
19	71.5	5.8	775	1	US-07-966-278-1
20	71.5	5.8	775	1	US-08-424-921-1
21	71.5	5.8	775	1	US-08-556-355A-1
22	71.5	5.8	775	2	US-07-803-627A-1
23	71.5	5.8	776	2	US-08-688-649-37
24	71.5	5.8	1114	2	US-08-576-626A-31
25	71	5.8	214	1	US-08-458-516-12
26	71	5.8	214	2	US-07-934-373C-39
27	71	5.8	214	3	US-08-437-642B-39
28	71	5.8	214	4	PCT-US93-07832-39

29	70	5.7	212	1	US-08-462-965A-2	Sequence 2, Appl
30	70	5.7	212	1	US-08-462-169B-21	Sequence 21, Appl
31	70	5.7	212	3	US-09-103-079-21	Sequence 21, Appl
32	70	5.7	214	1	US-08-425-763-1	Sequence 1, Appl
33	70	5.7	214	3	US-08-811-757-1	Sequence 1, Appl
34	70	5.7	1102	2	US-08-916-917-4	Sequence 4, Appl
35	70	5.7	1102	2	US-08-972-631-4	Sequence 4, Appl
36	70	5.7	1102	2	US-08-972-629-4	Sequence 4, Appl
37	70	5.7	1102	2	US-08-972-630-4	Sequence 4, Appl
38	70	5.7	1102	2	US-08-672-211-4	Sequence 4, Appl
39	70	5.7	1102	3	US-09-225-170-4	Sequence 4, Appl
40	69	5.6	214	2	US-07-934-373C-24	Sequence 24, Appl
41	69	5.6	214	3	US-08-437-642B-24	Sequence 24, Appl
42	69	5.6	214	4	PCT-US93-07832-24	Sequence 24, Appl
43	68.5	5.6	472	1	US-08-203-905B-14	Sequence 14, Appl
44	68.5	5.6	575	3	US-08-922-865-2	Sequence 2, Appl
45	68	5.6	474	2	US-08-282-197C-58	Sequence 58, Appl

ALIGNMENTS

RESULT 1
US-09-208-718-3
; Sequence 3, Application US/09208718
; Patent No. 6063767
; GENERAL INFORMATION:
; APPLICANT: Lai, Preeti
; APPLICANT: Hillman, Jennifer
; APPLICANT: Corley, Neil
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PHOSPHOLIPID BINDING PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/208,718
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/958,820
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0379 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 227 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LUNGUT12
; CLONE: 3126479
US-09-208-718-3

Query Match 99.1%; Score 1213; DB 3; Length 227;
Best Local Similarity 99.1%; Pred. No. 5.7e-131;

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)293-7060
TELEFAX: (202)293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-403-378B-4

Query Match 17.2%; Score 210.5; DB 1; Length 187;
Best Local Similarity 36.3%; Pred. No. 2.3e-16;
Matches 57; Conservative 20; Mismatches 59; Indels 21; Gaps 7;
QY 52 ELGNIGKVVPCNNYRQKITSWMEPIVKFPGAVDGATYILVMVDPDAPSRAPRQRFWR 111
Db 36 ELGKV---LTPQVMNRPSSISW-----DGLDPGKLYTLVLTDPDAPSRKDPKREWH 85
QY 112 HNLVTDIKGADLKKGTQGOELSAQAPSPPAHSGFH-RYQFFVYLOEGKVIS-----LLP 166
Db 86 HFLVNMKGNDISSGTV---LSEY-VGSGPKDTGLHRRYVWLVEQEQPLNCDEPILSN 140
QY 167 KENKTRGSKMDRFLNRFHLGPEASTQFMQYQDS 203
Db 141 KSGDNRGKFKVESFRKKYHLGAPVAGTCFOAE-WDDS 176

RESULT 4

US-08-403-378B-15
Sequence 15, Application US/08403378B
Patent No. 5759991

GENERAL INFORMATION:

APPLICANT: TOHDOH, NAOKI
APPLICANT: TOJO, SHIN-ICHIRO
APPLICANT: KOJIMA, SHIN-ICHI
APPLICANT: UEKI, YASUYUKI
APPLICANT: NISHIHARA, TOSHIO
APPLICANT: FUKUSHIMA, NOBUYOKI
APPLICANT: IRIE, TSUNEMASA
APPLICANT: ONO, KEIICHI
APPLICANT: AGUI, HIDEO
APPLICANT: OJIKI, KOSEI
TITLE OF INVENTION: NEUTROTROPIC PEPTIDE DERIVATIVES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:

ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 PENNSYLVANIA AVENUE, NW
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.

ZIP: 20037-3202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,378B
FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-124688
FILING DATE: 27-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1-080398
FILING DATE: 30-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1-280590
FILING DATE: 27-OCT-1989
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 1-333241
FILING DATE: 21-DEC-1989
PRIOR APPLICATION DATA: JP 2-243003
FILING DATE: 12-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/758,043
FILING DATE: 12-SEP-1991
PRIOR APPLICATION DATA: US 07/873,764
FILING DATE: 27-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/01214
FILING DATE: 27-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: BIGGART, WADDELL A
REGISTRATION NUMBER: 24,861
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)293-7060
TELEFAX: (202)293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-403-378B-15

Query Match 16.7%; Score 205; DB 1; Length 187;

Best Local Similarity 33.1%; Pred. No. 9.9e-16;

Matches 50; Conservative 21; Mismatches 62; Indels 18; Gaps 4;

QY 52 ELGNIGKVVPCNNYRQKITSWMEPIVKFPGAVDGATYILVMVDPDAPSRAPRQRFWR 111

Db 36 ELGKV---LTPQVKNRPTSISW-----DGLDSGKLYTLVLTDPDAPSRKDPKREWH 85

QY 112 HNLVTDIKGADLKKGTQGOELSAQAPSPPAHSGFHRYQFFVYLOEGKVIS-----LLPK 167

Db 86 HFLVNMKGNDISSGTV---LSDYVGGPKGTGLHRYVWLVEQDRPLKCDPEILSNR 141

QY 168 ENKTRGSKMDRFLNRFHLGPEASTQFMQ 198

Db 142 SGDHRGKFKVASFRKKYELRAPVAGTCYQAE 172

RESULT 5

US-07-644-372-2

Sequence 2, Application US/07644372

Patent No. 5416009

GENERAL INFORMATION:

APPLICANT: Lazzeri, Mario E.
APPLICANT: Nutman, Thomas B.
APPLICANT: Weiss, Niklaus
TITLE OF INVENTION: A DNA SEGMENT ENCODING A SPECIFIC
TITLE OF INVENTION: IMMUNODIAGNOSTIC ANTIGEN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1615 L. Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/644,372

Matches 45; **Conservative** 34; **Mismatches** 84; **Indels** 58; **Gaps** 10;

[illegible]

```

RESULT      8
US-07-934-373C-25
; Sequence 25, Application US/07934373C
; Patent No. 5821337
; GENERAL INFORMATION:
; APPLICANT: Paul J. Carter
; APPLICANT: Leonard G. Presta
; TITLE OF INVENTION: Immunglobulin Variants
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA

```

```

, ZIP: 34080
, COMPUTER READABLE FORM:
, MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
, MEDICAL INSTRUMENT: IBM PC compatible
, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: WinPatIn (Genentech)
,
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: US/07/934,373C
, FILING DATE: 21-Aug-1992
, CLASSIFICATION: 530
,
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: PCT/US92/05126
, FILING DATE: 15-JUN-1992
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: 07/715272
, FILING DATE: 14-JUN-1991
, ATTORNEY/AGENT INFORMATION:
, NAME: Lee, Wendy M.
, REGISTRATION NUMBER: 40,378
, REFERENCE/DOCKET NUMBER: P0709P2
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 650/225-1994
, TELEFAX: 650/952-9881
, INFORMATION FOR SEQ ID NO: 25:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 233 amino acids
, TYPE: Amino Acid
, TOPOLOGY: Linear
,
, US-07-934-373C-25

```

Query Match 6.2%; Score 76; DB 2; Length 233;
Best Local Similarity 23.5%; Pred. No. 0.79;
Matches 61; Conservative 30; Mismatches 89; Indels 80; Gaps 17;

QY 1 MGWTMLVTAALLGLMMVVTGDEDE----NSPCAHEALLDDEDTLFCQGLEVFYPELIGNI 56
||| : : | : || : || : | : | : |
Db 1 MGW-----SCIILFLVATATGVHSDIOMTQSPSLSASVGRV-----TI 40

Qy	57	GCKVVPDCNN	---YRQK-----	---TTSWMEPIV--	KFGCAVDGATYILVWVDDPAP	100
		:	:	:	:	
		:	:	:	:	
Db	41	TCRASQDINNLYNWOQKQKAPKPLLIYYTSTLHSGVPSRFGSGGTDTYTLTI	-----	---	---	94
		:	:	:	:	
		:	:	:	:	
Qy	101	SRAEPRQFRWHVLVTDIKGADLKKGKIQOELS	---AYQAPS---	PAHSGFHYQFF	153	
		:	:	:	:	
		:	:	:	:	
Db	95	SSQQP-EDFATYCYQ	---QNGTLPPFGQGTQKVEIKRTVAAPSVFIFPPSDE	-----	142	
		:	:	:	:	
		:	:	:	:	
Qy	154	VYLQEG--KVISLT	---PRENKTRGSKMDRFLNRFHLCPEASTQFMTQNYQDSPTLQ	207		
		:	:	:	:	
		:	:	:	:	
Db	143	QLKSGTASVVCVLLNNPYPREAKVQ	--WKVDNALQSGNSQESVTEQDSKDSYLSLSSTLT	199		
		:	:	:	:	
		:	:	:	:	
Qy	208	APRERASEPKHKNOAEIAC	227			
		:	:	:	:	
		:	:	:	:	
Db	200	L--SKADYERKHK	-----VIAC	213		

```

RESULT          9
US-08-437-642B-25
; Sequence 25, Application US/08437642B
; Patent No. 6054297
; GENERAL INFORMATION:
; APPLICANT: Paul J. Carter
; APPLICANT: Leonard G. Presta
; TITLE OF INVENTION: Immunoglobulin Variants
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080

```

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/437,642B
FILING DATE: 09-May-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: 07/934373
FILING DATE: 21-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/146206
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/05126
FILING DATE: 15-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/715272
FILING DATE: 14-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 40,378
REFERENCE/DOCKET NUMBER: P0709P2C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1994
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 233 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-437-642B-25

Query Match 6.2%; Score 76; DB 3; Length 233;
Best Local Similarity 23.5%; Pred. No. 0.79;
Matches 61; Conservative 30; Mismatches 89; Indels

[illegible]

RESULT 11
US-08-916-770-3
; Sequence 3, Application US/08916770

1 PATENT NO.: 3674248
 2
 3 GENERAL INFORMATION:
 4 APPLICANT: Hillman, Jennifer
 5 APPLICANT: Corley, Neil C.
 6 APPLICANT: Shah, Purvi
 7 TITLE OF INVENTION: GLUTATHIONE S-TRANSFERASE
 8 TITLE OF INVENTION: HOMOLOG
 9 NUMBER OF SEQUENCES: 3
 10
 11 CORRESPONDENCE ADDRESS:
 12 ADDRESSEE: Incyte Pharmaceuticals, Inc.

ADDRESSES: 3174 Porter Dr.
STREET: 3174 Porter Dr.
CITY: Palo Alto

STATE: CA
COUNTRY: USA
ZIP: 94304

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
;
```

```
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/08/916,770
; FILING DATE: Filed Herewith
; REPT. INFORMATION DATA:

```

```

; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER:
;
; FILING DATE:
;

```

FILED DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0370 US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:

```
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 212 amino acids
;     TYPE:   amino acid
```

```

; FILE, QMNO FILE
; STRANDEDNESS: single
; TOPOLOGY: linear
;

```

```

;
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
;

```

CLONE: 1750200
US-08-916-770-3

603

Query Match: 0.00;
Best Local Similarity: 23.28;
Matches: 45: Conservative: 3

CONCLUSIONS

QY	1	MGWTRLVTAALLGLMMVVTGDEDE---	NSPCAHEALLDEDTLFCGLEVFYFELGNI	56
		: : : :		
Db	1	MGW-----SCIIELVATATGVHSDIOMTQVSPSSLSASGDRV---		40
		: : : :		
QY	57	GKVVPPDCNN---YRQK-----ITSMFEIV--KFGAVDGTATYILVMVDDPAP	100	
		: : : :		
Db	41	TCRASQDINNYLNWYQOKPGKAPKLIITYSTLTHSGVPSRFGSGSGDYTLTI-----	94	
		: : : :		
QY	101	SRAEPQRERWHLVTDIDKADLKKGIQOGLS---AYQAPS---PPAISGFHRYOFF	153	
		: : : :		
Db	95	SSLQP--EDFATYYCQ--QGNTLPTFFGGTKVEIKRTVAAPSVFIFFPSDE-----	142	
		: : : :		
QY	154	VYLQEG--KVISLL-----PKENKTRGSWKMDRFLNRFHLGEPEASTQPMWQNOVDSPTLQ	207	
		: : : :		
Db	143	-QLKSGTASVWCLLNIFYPREAKVQ--WKVDNALQSGNSQSVTEQDSKDSITYLSLSHLT	199	
		: : : :		
QY	208	APRERASEPKHKNQAEIAAC	227	
		: : :		
Db	200	L--SKADYEKKH-----VYAC	213	
		: : :		

RESULT 10
PCT-US93-07832-25
; Sequence 25, Application PC/TUS9307832

```

: GENERAL INFORMATION:
: APPLICANT: Genentech, Inc.
: TITLE OF INVENTION: Immunglobulin Variants
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
:

```

COUNTRY: USA
ZIP: 94080

```

;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS 4.0/MS-DOS

```

```

: OPERATING SYSTEM PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/07832

```

APPLICATION NUMBER: FC190555/01032
FILING DATE: 19930820
CLASSIFICATION:

PRIOR APPLICATION DATA: 07/715272
APPLICATION NUMBER: 07/715272

```

; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: PCT/US92/05126
FILING DATE: 15-JUN-1992

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/934373
 ; FILING DATE: 21-AUG-1992
 ;

FILED DATE: 21 NOV 1992
; ATTORNEY/AGENT INFORMATION:
; NAME:

REGISTRATION NUMBER: 709P2PCT
REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:
TELEPHONE:

TELEFAX: 415/952-9881
TELEX: 910/371-7168

```

; INFORMATION FOR SEQ ID NO: 23:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 233 amino acids

```

LEUCIN: 200 amino acids
TYPE: amino acid
TOPOLOGY: linear

PCT-US93-07832-25

Query Match	6.28; Score 76
	02 58

Best Local Similarity 23.5%; Fred. No.
Matches 61; Conservative 30; Mismatch

Db 533 PTPDEGDRVRAEMPQLRKQLE 555

Search completed: March 24, 2001, 13:23:55
Job time: 145 sec

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	147	22.1	112	1	COL_CANFA	P19090	canis famli
2	141	21.2	107	1	COL_RABIT	P42890	oryctolagus
3	135.5	20.3	112	1	COL_HUMAN	P04118	homo sapien
4	134	20.1	95	1	COL2_PIG	P02703	sus scrofa
5	134	20.1	108	1	COLB_HORSE	P02705	equus caball
6	131	19.7	106	1	COLA_HORSE	P02704	equus cabal
7	131	19.7	111	1	COL_MYOCO	P42889	myocastor c
8	122	18.3	112	1	COL_RAT	P17084	rattus norv
9	79	11.9	226	1	PDGA_XENIA	P13698	xenopus lae
10	78.5	11.8	2531	1	NTCL1_RAT	Q07008	rattus norv
11	78	11.7	96	1	RS1L_LYCRES	P47926	lycopersico
12	76	11.4	2139	1	CRB_DROME	P10040	dtrosophila
13	75.5	11.3	2531	1	NTCL1_MOUSE	Q01705	mus musculu
14	75	11.3	383	1	DLK_HUMAN	P80370	homo sapien
15	74.5	11.2	571	1	DTSJ3_BOTYA	P30431	bothrops ja
16	74.5	11.2	1700	1	BAR3_CHITE	Q03376	chironomus
17	74	11.1	385	1	DLK_MOUSE	Q09163	mus musculu
18	73	11.0	112	1	GST1_LYCRES	P27057	lycopersico
19	73	11.0	806	1	IFB7_MOUSE	P26011	mus musculu
20	73	11.0	807	1	FSPQ_RAT	P35446	rattus norv
21	72	10.8	211	1	PDGA_HUMAN	P04085	homo sapien
22	72	10.8	1221	1	FB12_MOUSE	P37889	mus musculu
23	71.5	10.7	81	1	VPRA_DENPO	P25687	dendroaspis
24	71.5	10.7	369	1	DNAT1_LEPIN	P71443	leptospiira
25	71.5	10.7	1790	1	LMB1_DROME	P11046	dtrosophila
26	71	10.7	211	1	PDGA_MOUSE	P20033	mus musculu
27	71	10.7	2907	1	FBN2_MOUSE	Q61555	mus musculu
28	71	10.7	2911	1	FBN2_HUMAN	P35556	homo sapien
29	70.5	10.6	1323	1	LT23_CAEL	P24348	caenorhabdi
30	70	10.5	798	1	IFB7_HUMAN	P26010	homo sapien
31	70	10.5	1895	1	YLK3_CAEL	P41951	caenorhabdi
32	70	10.5	2703	1	NOTPC_DROME	P07207	dtrosophila
33	69.5	10.4	78	1	THG_PETIN	Q40901	petunia int

```

SEQUENCE 107 AA; 11271 MW; 825BA1AEB1422390 CRC64;

Query Match 21.2%; Score 141; DB 1; Length 107;
Best Local Similarity 42.4%; Pred. NO. 4.5e-07;
Matches 25; Conservative 7; Mismatches 27; Indels 0; Gaps 0;

QY 33 LELKESIRNQDCTGCCORAPDNCESHCAEKSEGLCOTQVFFGQYRACPLRLNLTG 91
  || || || || || || || || || || || || || || || || || || || ||
DBD 28 LEEGELCLNSAQKSCGCHHSSALSARCAPKASENSECFQIYGVYKCPGERGLTC 86

RESULT 3
COL_HUMAN COL_HUMAN STANDARD; PRT; 112 AA.
ID COL_HUMAN
AC P04118;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE COLIPASE PRECURSOR.
DE CLPS
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
RN
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=90248429; PubMed=2337598;
RA Lowe M.E., Rosenblum J.L., McEwen P., Strauss A.W.;
RT "Cloning and characterization of the human colipase cDNA.";
RL Biochemistry 29:823-828(1990).
[2]
RN
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92353041; PubMed=1643046;
RA Sims H.F., Lowe M.E.;
RT "The human colipase gene: isolation, chromosomal location, and
  tissue-specific expression".
RL Biochemistry 31:7120-7125(1992).
[3]
RN
RN
RP SEQUENCE OF 23-108.
RX MEDLINE=84104937; PubMed=6691986;
RA Sternby B., Engstroem A., Hellman U., Viht A.M., Sternby N.-H.,
  Borgstroem B.;
RT "The primary sequence of human pancreatic colipase.";
RL Biochim. Biophys. Acta 784:75-80(1984).
[4]
RN
RN
RX X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=93241293; PubMed=8479519;
RA van Tilbeurgh H., Egloff M.-P., Martinez C., Rugani N., Verger R.,
  Cambillau C.;
RT "Interfacial activation of the lipase-procolipase complex by mixed
  micelles revealed by X-ray crystallography.";
RL Nature 362:814-820(1993).
CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
  THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
  COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
  INHIBITORY EFFECT ON THE LIPASE.
CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
  SIGNAL.
CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation in
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announcement/
  or send an email to license@isb-sib.ch).
CC
CC EMBL: J02883; AAA52054.1; -.
CC EMBL: M95529; AAB05818.1; -.
CC PIR: A03163; XLHU.
CC PIR: A33949; A33949.
CC PIR: A42568; A42568.

```


DR HSP; P02703; 1ETH.
 DR MIM; 120105; -
 DR INTERPRO; IPR001981; -
 DR PFAM; PF01114; Colipase; 1.
 DR PRINTS; PR00128; COLIPASE.
 DR PROSITE; PS00121; COLIPASE; 1.
 KW Lipid degradation; Digestion; Pancreas; Signal.
 FT SIGNAL 1 17
 FT PROPEP 18 22
 FT ENTEROSTATIN, ACTIVATION PEPTIDE
 FT (POTENTIAL).
 FT CHAIN 23 112
 FT DISULFID 34 45
 FT DISULFID 40 56
 FT DISULFID 44 78
 FT DISULFID 66 86
 FT DISULFID 80 104
 FT CONFLICT 68 69
 FT MISSING (IN REF. 2).
 SQ SEQUENCE 112 AA; 11954 MW; 772872EBBE7C4DF8 CRC64;
 Query Match 20.3%; Score 135.5; DB 1; Length 112;
 Best Local Similarity 28.8%; Pred. No. 1.6e-06;
 Matches 32; Conservative 17; Mismatches 51; Indels 11; Gaps 3;
 QY 12 LFFFLFLTRGSLPTKYNL---LEKESCIRNQDCETGCCQAPDNCSEHCAKRGSEG 68
 :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
 Db 4 ILILLVALSVAVAPGPRGIILINLENGELCMNSACKSNCCQHSSALGLARCTSMASEN 63
 QY 69 SLQQTQVFFQYACACLRNLTCIYSKNEKWL-----SIAYGRCQKIGRK 114
 :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
 Db 64 SEGSVRLTYGIYKPCERGLTC---EGDKTVGSITNTNFGICHVGRSK 111
 RESULT 4
 COL2_PIG STANDARD; PRT; 95 AA.
 AC P02703;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE COLIPASE II PRECURSOR (PROCOLIPASE II).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE.
 RX MEDLINE-84104937; PubMed-6691986;
 RA Sternby B., Engstroem A., Hellman U., Vihert A.M., Sternby N.-H.,
 RA Borgstroem B.;
 RL "The primary sequence of human pancreatic colipase.";
 RL Biochim. Biophys. Acta 784:75-80(1984).
 RN [2]
 RP SEQUENCE OF 6-91.
 RX MEDLINE-74290109; PubMed-4858821;
 RA Charles M., Erlanson C., Blanchetta J.D., Joffre J., Guidoni A.A.,
 RA Ravery M.;
 RL "The primary structure of porcine colipase II. I. The amino acid
 RL sequence.";
 RL Biochim. Biophys. Acta 359:186-197(1974).
 RN [3]
 RP DISULFIDE BONDS.
 RX MEDLINE-74290110; PubMed-4603223;
 RA Erlanson C., Charles M., Astier M., Desnuelle P.;
 RL "The primary structure of porcine colipase II. II. The disulfide
 RL bridges.";
 RL Biochim. Biophys. Acta 359:198-203(1974).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RX MEDLINE-92396238; PubMed-1522902;
 RA van Tilbeurgh H., Sarda L., Verger R., Cambillau C.;
 RL "Structure of the pancreatic lipase-procolipase complex.";
 RN Nature 359:159-162(1992).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE-95291181; PubMed-7773176;
 RA Egloff M.-P., Sarda L., Verger R., Cambillau C., van Tilbeurgh H.;
 RL "Crystallographic study of the structure of colipase and of the
 RL interaction with pancreatic lipase.";
 RL Protein Sci. 4:44-57(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE-96279347; PubMed-8663362;
 RA Hermoso J., Pignol D., Kerfelec B., Crenon I., Chapus C.,
 RA Fontecilla-Camps J.C.;
 RL "Lipase activation by nonionic detergents. The crystal structure of
 RL the porcine lipase-colipase-tetraethylene glycol monoethyl ether
 RL complex.";
 RL J. Biol. Chem. 271:18007-18016(1996).
 RN [7]
 RP STRUCTURE BY NMR, AND REVISION TO 37.
 RX MEDLINE-95172049; PubMed-7867624;
 RA Breg J.N., Sarda L., Cozzone P.J., Rugani N., Boelens R., Kaptein R.;
 RL "Solution structure of porcine pancreatic procolipase as determined
 RL from 1H homonuclear two-dimensional and three-dimensional NMR.";
 RL Eur. J. Biochem. 227:663-672(1995).
 CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
 CC THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
 CC COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
 CC INHIBITORY EFFECT ON THE LIPASE.
 CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
 CC SIGNAL.
 CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
 CC PIR; A03162; XLP62.
 DR PDB; 1PCN; 20-DEC-94.
 DR PDB; 1PCO; 20-DEC-94.
 DR PDB; 1LPA; 01-NOV-94.
 DR PDB; 1LPB; 01-NOV-94.
 DR PDB; 1ETH; 07-DEC-96.
 DR INTERPRO; IPR001981; -
 DR PFAM; PF01114; Colipase; 1.
 DR PRINTS; PR00128; COLIPASE.
 DR PROSITE; PS00121; COLIPASE; 1.
 KW Lipid degradation; Digestion; Pancreas; 3D-structure.
 FT PROPEP 1 5
 FT CHAIN 6 95
 FT SITE 6 9
 FT ENTEROSTATIN, ACTIVATION PEPTIDE.
 FT ESSENTIAL FOR THE FUNCTION OF CLP IN THE
 FT PRESENCE OF PHOSPHOLIPIDS.
 FT ESSENTIAL FOR THE FUNCTION OF CLP IN THE
 FT PRESENCE OF PHOSPHOLIPIDS.
 FT ACT_SITE 92 92
 FT DISULFID 17 28
 FT DISULFID 23 39
 FT DISULFID 27 61
 FT DISULFID 49 69
 FT DISULFID 63 87
 FT CONFLICT 37 37
 FT MISSING (IN REF. 2).
 FT CONFLICT 50 50
 FT MISSING (IN REF. 2).
 SQ SEQUENCE 95 AA; 10309 MW; 423E04549E196027 CRC64;
 Query Match 20.1%; Score 134; DB 1; Length 95;
 Best Local Similarity 34.1%; Pred. No. 1.9e-06;
 Matches 29; Conservative 11; Mismatches 37; Indels 8; Gaps 2;
 QY 33 LELKESCIRNQDCETGCCQAPDNCSEHCAKRGSEGSLQQTQVFFQYACACLRNLTCI 92
 :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
 Db 11 LDEGELCLNSAQCKSNCCQHDHTILSRCAKARENSECSAFTLYGYKPCERGLTC- 69
 QY 93 YSKNEKWL-----SIAYGRCQKIGR 112
 :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
 Db 70 --EGDKSLVGSITNTNFGICHVGR 92
 RESULT 5
 COLB_HORSE STANDARD; PRT; 108 AA.
 ID COLB_HORSE
 AC P02705;

DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PROCOLIPASE B PRECURSOR (FRAGMENT)
OS Equus caballus (Horse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
RN [1]
RN SEQUENCE FROM N.A.
RN MEDLINE=82046794; PubMed=7295770;
RN Bonicel J.J., Couchoud P.M., Foglizzo E., Desnuelle P., Chapus C.;
RN "Amino acid sequence of horse colipase B";
RN Biochim. Biophys. Acta 669:39-45(1981).
RN [3]
RN SEQUENCE OF 14-68.
RN MEDLINE=81021166; PubMed=7417313;
RN Julien R., Bechis G., Gregoire J., Rathelot J., Rochat H., Sarda L.;
RN "Evidence for the existence of two isocolipases in horse pancreas";
RN Biochem. Biophys. Res. Commun. 95:1245-1252(1980).
RN -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
RN THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
RN COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
RN INHIBITORY EFFECT ON THE LIPASE.
RN -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
RN SIGNAL.
RN -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RN or send an email to license@isb-sib.ch).
RN EMBL; X74344; CAA52391.1;
RN PIR; A03165; XLHOB.
RN HSSP; P02703; IETH.
RN INTERPRO; IPR001981;
RN PFAM; PF01114; Colipase; 1.
RN PROSITE; PS00121; COLIPASE; 1.
RN Lipid degradation; Digestion; Pancreas; Signal.
RN NON_TER 1
RN SIGNAL 1
RN PROPEP 13
RN CHAIN 18
RN BINDING 19
RN BINDING 65
RN DISULFID 30
RN DISULFID 41
RN DISULFID 36
RN DISULFID 52
RN DISULFID 40
RN DISULFID 74
RN DISULFID 62
RN DISULFID 82
RN DISULFID 76
RN DISULFID 100
RN CONFLICT 35
RN CONFLICT 35
RN CONFLICT 42
RN CONFLICT 42
RN CONFLICT 108
RN CONFLICT 108
RN SEQUENCE 108 AA; 11618 MW; 37A7F49ABA08B30 CRC64;
Query Match 20.1%; Score 134; DB 1; Length 108;
Best Local Similarity 39.7%; Pred. No. 2.1e-06;
Matches 25; Conservative 6; Mismatches 32; Indels 0; Gaps 0;
OY 33 LEKESCIRNQDCGCGCORAFDNCSEHCAEKGSGSLCQTVFFGOYRACPCRLNLCI 92
DB 24 LEAGEICMNSAQCKBCCCHRESSLSLARCACAKASECSAWTLGVYKPCPERGLTCQ 83

OY 93 YSK 95
DB 84 VDK 86
RESULT 6
COLA_HORSE STANDARD; PRT; 106 AA.
ID P02704;
AC 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PROCOLIPASE A PRECURSOR (FRAGMENT)
OS Equus caballus (Horse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
RN [1]
RN SEQUENCE FROM N.A.
RN TISSUE=PANCREAS;
RN MEDLINE=94325330; PubMed=8049250;
RN Crenon I., Granon S., Chapus C., Kerfelec B.;
RN "Molecular cloning and expression of two horse pancreatic CDNA
RN encoding colipase A and B";
RN Biochim. Biophys. Acta 1213:357-360(1994).
RN [2]
RN SEQUENCE OF 12-106.
RN MEDLINE=84104937; PubMed=6691986;
RN Sternby B., Engstroem A., Hellman U., Vihert A.M., Sternby N.-H.,
RN Borgstroem B.;
RN "The primary sequence of human pancreatic colipase";
RN Biochim. Biophys. Acta 784:75-80(1984).
RN [3]
RN SEQUENCE OF 12-106.
RN MEDLINE=82186702; PubMed=7075593;
RN Pierrot M., Astier J.-P., Astier M., Charles M., Drenth J.;
RN "Pancreatic colipase: crystallographic and biochemical aspects";
RN Eur. J. Biochem. 123:347-354(1982).
RN [4]
RN SEQUENCE OF 12-66.
RN MEDLINE=81021166; PubMed=7417313;
RN Julien R., Bechis G., Gregoire J., Rathelot J., Rochat H., Sarda L.;
RN "Evidence for the existence of two isocolipases in horse pancreas";
RN Biochem. Biophys. Res. Commun. 95:1245-1252(1980).
RN -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
RN THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
RN COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
RN INHIBITORY EFFECT ON THE LIPASE.
RN -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
RN SIGNAL.
RN -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RN or send an email to license@isb-sib.ch).
RN EMBL; X74503; CAA52611.1;
RN PIR; A03164; XLHOA.
RN HSSP; P02703; IETH.
RN INTERPRO; IPR001981;
RN PFAM; PF01114; Colipase; 1.
RN PROSITE; PS00121; COLIPASE; 1.
RN Lipid degradation; Digestion; Pancreas; Signal.
RN NON_TER 1
RN SIGNAL 1
RN PROPEP 11
RN CHAIN 12
RN BINDING 16
RN DISULFID 17
RN DISULFID 106
RN DISULFID 28
RN DISULFID 39
RN DISULFID 34
RN DISULFID 50
ENTEROSTATIN, ACTIVATION PEPTIDE.
COLIPASE A.
BY SIMILARITY.
BY SIMILARITY.

```

FT CHAIN      23    111 COLIPASE.
FT FT         23    45 BY SIMILARITY.
FT DISULFID   34    45 BY SIMILARITY.
FT FT         40    56 BY SIMILARITY.
FT DISULFID   44    78 BY SIMILARITY.
FT FT         66    86 BY SIMILARITY.
FT DISULFID   80    104 BY SIMILARITY.
SQ SEQUENCE   111 AA; 11899 MW; 7FF37DE1C169616B CRC64;

Query Match          19.7%; Score 131; DB 1; Length 111;
Best Local Similarity 35.9%; Pred.No. 4.le-06;
Matches 28; Conservative 10; Mismatches 38; Indels 2; Gaps 1;

QY 33 LELKESIRNQDCETGCCQRAPDNCESHCAEKSEGSICQTQVFEGGYRACPCPLRNITCI 92
   : - - - : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 28 LDNGELCLNSAQCKSCQCQHDSPLGLARCADKARENSGCSPTQTIYIYLCP CERGLTCD 87

QY 93 YSKN--EKWLSIAYGRQC 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 88 GDKSIIGAITNTNYGICQ 105

RESULT 8
COL_RAT ID COL_RAT STANDARD; PRT; 112 AA.
AC PI7084;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE COLIPASE PRECURSOR.
DE CLPS.
GN Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=90179738; PubMed=2129524;
RA Wicker C., Puigserver A.;
RT "Rat pancreatic colipase mRNA: nucleotide sequence of a cDNA clone
and nutritional regulation by a lipidic diet";
RL Biochem. Biophys. Res. Commun. 167:130-136(1990).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=94262798; PubMed=8203536;
RA Payne R.M., Sims H.F., Jennens M.L., Lowe M.E.;
RT "Rat pancreatic lipase and two related proteins: enzymatic properties
and mRNA expression during development.";
RL Am. J. Physiol. 266:G914-G921(1994).
CC CC
CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
THE LIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
INHIBITORY EFFECT ON THE LIPASE.
CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
SIGNAL.
CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
of the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/ or
send an email to license@isb-sib.ch).
CC -----
CC EMBL; M33333; AAA40943.1; -.
CC DR EMBL; M58370; AAA20505.1; -.
CC DR PIR; A34623; A34623.
CC DR HSSP; P02703; IETH.
CC DR INTERPRO; IPR001981; -.
CC DR PFAM; PF01114; Colipase; 1.
CC DR PRINTS; PR00128; COLIPASE.
CC DR PROSITE; PS00121; COLIPASE; 1.
CC DR Lipid degradation; Digestion; Pancreas; Signal.
KW

```

SIGNAL	1	17
PROPEP	18	22
ENTEROSTATIN, ACTIVATION PEPTIDE (POTENTIAL). COLIPASE.		
CHAIN	23	112
DISULFID	34	45
BY SIMILARITY.	40	56
DISULFID	44	78
BY SIMILARITY.	44	78
DISULFID	66	86
BY SIMILARITY.	66	86
DISULFID	80	104
CONFLICT	18	18
SEQUENCE	112 AA;	12280 MW; FB5659DA59D04D9D CRC64;

Query Match 18.3%; Score 122; DB 1; Length 112;
Best Local Similarity 31.9%; Pred No. 3e-05; Matches 29; Conservative 8; Mismatches 38; Indels 16; Gaps 2;

OY	33	LELKESCIRNPQCTGCCORAPDNCEHCAKSGESGLCQTQVFGQYRACPLRLNLTG- 91
Ddb	28	LDEGEICVNSMOCKSRCCQHDITLGIARTCHRAMENSECSPKTLGYIYYRCPCERGLTCE 87
OY	92	-----LYSKNEKWLSIAYGRCKQKIGROK 114
Ddb	88	GDRSIIGAINTNTN-----YGVLDSSTRSK 111

RESULT 9
PDGA_XENLA
ID PDGA_XENLA STANDARD; PRT; 226 AA.
AC PI3698;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PLATELET-DERIVED GROWTH FACTOR, A CHAIN PRECURSOR (PDGF A-CHAIN) (PDGFA).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=OOCYTE;
RC MEDLINE=88321676; PubMed=3413486;
RX Mercola M., Melton D.A., Stiles C.D.;
RT "platelet-derived growth factor A chain is maternally encoded in
Xenopus embryos.";
RL Science 241:1223-1225(1988).
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE=OOCYTE;
RX MEDLINE=90175018; PubMed=2308861;
RA Bejcek B.E., Li D.Y., Deuel T.F.;
RT "Nucleotide sequence of a cDNA clone of Xenopus platelet-derived
growth factor A-chain.";
RL Nucleic Acids Res. 18:680-680(1990).
CC -1- FUNCTION: PLATELET-DERIVED GROWTH FACTOR IS A POTENT MITOGEN FOR
CELLS OF MESENCHYMAL ORIGIN. BINDING OF THIS GROWTH FACTOR TO ITS
AFFINITY RECEPTOR ELICITS A VARIETY OF CELLULAR RESPONSES. IT IS
RELEASED BY PLATELETS UPON WOUNDING AND PLAYS AN IMPORTANT ROLE
IN STIMULATING ADJACENT CELLS TO GROW AND THEREBY HEAL THE WOUND.
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED DIMER OF NONIDENTICAL (A
AND B) CHAINS. HOMODIMERS OF A AND B CHAINS ARE IMPLICATED IN
TRANSFORMATION PROCESSES.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE LONG FORM
CONTAINS A BASIC INSERT WHICH ACTS AS A CELL RETENTION SIGNAL.
CC -1- MISCELLANEOUS: A-A AND B-B, AS WELL AS A-B, DIMERS CAN BIND TO THE
PDGF RECEPTOR.
CC -1- SIMILARITY: BELONGS TO THE PDGF/VEGF FAMILY OF GROWTH FACTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation at
the European Bioinformatics Institute. There are no restrictions on its

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
 CC DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
 CC ADULT.
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X57405; CAA040667.1; -
 CC HSSP: P00740; LIXA.
 CC INTERPRO: IPR000152; -
 CC INTERPRO: IPR000561; -
 CC INTERPRO: IPR000800; -
 CC INTERPRO: IPR001438; -
 CC INTERPRO: IPR001881; -
 CC INTERPRO: IPR002049; -
 CC INTERPRO: IPR002110; -
 CC PFAM: PF00008; EGF_36.
 CC PFAM: PF00023; ank; 6.
 CC PFAM: PF00066; notch; 3.
 CC PRINTS: PR00010; EGF-BLOOD.
 CC PRINTS: PR00011; EGF-LAMININ.
 CC PROSITE: PS00088; ANK_REPEAT; 4.
 CC PROSITE: PS0297; ANK_REPEAT_REGION; 1.
 CC PROSITE: PS00010; ASX_HYDROXYL; 22.
 CC PROSITE: PS00022; EGF_1; 35.
 CC PROSITE: PS01186; EGF_2; 26.
 CC PROSITE: PS01187; EGF_CA; 21.
 CC Differentiation: Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 FT DOMAIN 19 1723 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1724 1746 POTENTIAL.
 FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 20 58 EGF-LIKE 1.
 FT DOMAIN 59 99 EGF-LIKE 2.
 FT DOMAIN 102 139 EGF-LIKE 3.
 FT DOMAIN 140 176 EGF-LIKE 4.
 FT DOMAIN 178 216 EGF-LIKE 5.
 FT DOMAIN 218 255 EGF-LIKE 6.
 FT DOMAIN 257 293 EGF-LIKE 7.
 FT DOMAIN 295 333 EGF-LIKE 8.
 FT DOMAIN 335 371 EGF-LIKE 9.
 FT DOMAIN 372 410 EGF-LIKE 10.
 FT DOMAIN 412 450 EGF-LIKE 11.
 FT DOMAIN 452 488 EGF-LIKE 12.
 FT DOMAIN 490 526 EGF-LIKE 13.
 FT DOMAIN 528 564 EGF-LIKE 14.
 FT DOMAIN 566 601 EGF-LIKE 15.
 FT DOMAIN 603 639 EGF-LIKE 16.
 FT DOMAIN 641 676 EGF-LIKE 17.
 FT DOMAIN 678 714 EGF-LIKE 18.
 FT DOMAIN 716 751 EGF-LIKE 19.
 FT DOMAIN 753 789 EGF-LIKE 20.
 FT DOMAIN 791 827 EGF-LIKE 21.
 FT DOMAIN 829 867 EGF-LIKE 22.
 FT DOMAIN 869 905 EGF-LIKE 23.
 FT DOMAIN 907 943 EGF-LIKE 24.
 FT DOMAIN 945 981 EGF-LIKE 25.
 FT DOMAIN 983 1019 EGF-LIKE 26.
 FT DOMAIN 1021 1057 EGF-LIKE 27.
 FT DOMAIN 1059 1095 EGF-LIKE 28.
 FT DOMAIN 1097 1143 EGF-LIKE 29.

FT DOMAIN 1145 1181 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1183 1219 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1221 1265 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1267 1305 EGF-LIKE 33.
 FT DOMAIN 1307 1346 EGF-LIKE 34.
 FT DOMAIN 1348 1384 EGF-LIKE 35.
 FT DOMAIN 1387 1426 EGF-LIKE 36.
 FT DOMAIN 1449 1462 CYS-RICH.
 FT DOMAIN 1865 2076 6 X ANK MOTIF REPEATS.
 FT REPEAT 1865 1910 ANK MOTIF 1.
 FT REPEAT 1912 1942 ANK MOTIF 2.
 FT REPEAT 1944 1975 ANK MOTIF 3.
 FT REPEAT 1978 2009 ANK MOTIF 4.
 FT REPEAT 2011 2042 ANK MOTIF 5.
 FT REPEAT 2044 2076 ANK MOTIF 6.
 FT DISULFID 24 37 BY SIMILARITY.
 FT DISULFID 31 46 BY SIMILARITY.
 FT DISULFID 48 57 BY SIMILARITY.
 FT DISULFID 63 74 BY SIMILARITY.
 FT DISULFID 68 87 BY SIMILARITY.
 FT DISULFID 89 98 BY SIMILARITY.
 FT DISULFID 106 117 BY SIMILARITY.
 FT DISULFID 111 127 BY SIMILARITY.
 FT DISULFID 129 138 BY SIMILARITY.
 FT DISULFID 144 155 BY SIMILARITY.
 FT DISULFID 149 164 BY SIMILARITY.
 FT DISULFID 166 175 BY SIMILARITY.
 FT DISULFID 182 195 BY SIMILARITY.
 FT DISULFID 189 204 BY SIMILARITY.
 FT DISULFID 206 215 BY SIMILARITY.
 FT DISULFID 222 233 BY SIMILARITY.
 FT DISULFID 227 243 BY SIMILARITY.
 FT DISULFID 245 254 BY SIMILARITY.
 FT DISULFID 261 272 BY SIMILARITY.
 FT DISULFID 266 281 BY SIMILARITY.
 FT DISULFID 283 292 BY SIMILARITY.
 FT DISULFID 299 312 BY SIMILARITY.
 FT DISULFID 306 321 BY SIMILARITY.
 FT DISULFID 323 332 BY SIMILARITY.
 FT DISULFID 339 350 BY SIMILARITY.
 FT DISULFID 344 359 BY SIMILARITY.
 FT DISULFID 361 370 BY SIMILARITY.
 FT DISULFID 376 387 BY SIMILARITY.
 FT DISULFID 381 398 BY SIMILARITY.
 FT DISULFID 400 409 BY SIMILARITY.
 FT DISULFID 416 429 BY SIMILARITY.
 FT DISULFID 423 438 BY SIMILARITY.
 FT DISULFID 440 449 BY SIMILARITY.
 FT DISULFID 456 467 BY SIMILARITY.
 FT DISULFID 461 476 BY SIMILARITY.
 FT DISULFID 478 487 BY SIMILARITY.
 FT DISULFID 494 505 BY SIMILARITY.
 FT DISULFID 499 514 BY SIMILARITY.
 FT DISULFID 516 525 BY SIMILARITY.
 FT DISULFID 532 543 BY SIMILARITY.
 FT DISULFID 537 552 BY SIMILARITY.
 FT DISULFID 554 563 BY SIMILARITY.
 FT DISULFID 570 580 BY SIMILARITY.
 FT DISULFID 575 589 BY SIMILARITY.
 FT DISULFID 591 600 BY SIMILARITY.
 FT DISULFID 607 618 BY SIMILARITY.
 FT DISULFID 612 627 BY SIMILARITY.
 FT DISULFID 629 638 BY SIMILARITY.
 FT DISULFID 645 655 BY SIMILARITY.
 FT DISULFID 650 664 BY SIMILARITY.
 FT DISULFID 666 675 BY SIMILARITY.
 FT DISULFID 682 693 BY SIMILARITY.
 FT DISULFID 687 702 BY SIMILARITY.
 FT DISULFID 704 713 BY SIMILARITY.
 FT DISULFID 720 730 BY SIMILARITY.
 FT DISULFID 725 739 BY SIMILARITY.
 FT DISULFID 741 750 BY SIMILARITY.
 FT DISULFID 757 768 BY SIMILARITY.

[illegible]

```
CC -----
DR EMBL; M33753; AAA28428.1; ALT_SEQ.
DR EMBL; X05144; CAA28793.1; -.
DR PIR; B26637; B26637.
DR PIR; A36672; A36672.
DR HSSP; P00740; IIXA.
DR FLYBASE; FBgn0000368; crb.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR PFAM; PF00008; EGF; 27.
DR PFAM; PF00054; laminin_G; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PROSITE; PS00010; ASX_HYDROXYL; 15.
DR PROSITE; PS00022; EGF_1; 26.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS01187; EGF_CA; 15.
KW Differentiation; Repeat; EGF-like domain; Transmembrane;
KW Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 90
FT CHAIN 91 2139
FT DOMAIN 91 2139 CRUMBS PROTEIN.
FT TRANSMEM 91 2084 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2085 2111 POTENTIAL.
FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 267 303
FT DOMAIN 306 343 EGF-LIKE 1.
FT DOMAIN 348 386 EGF-LIKE 2.
FT DOMAIN 388 425 EGF-LIKE 3.
FT DOMAIN 427 463 EGF-LIKE 4.
FT DOMAIN 464 500 EGF-LIKE 5.
FT DOMAIN 501 532 EGF-LIKE 6.
FT DOMAIN 545 581 EGF-LIKE 7.
FT DOMAIN 582 611 EGF-LIKE 8.
FT DOMAIN 609 646 EGF-LIKE 9.
FT DOMAIN 648 685 EGF-LIKE 10.
FT DOMAIN 687 723 EGF-LIKE 11.
FT DOMAIN 725 761 EGF-LIKE 12.
FT DOMAIN 763 800 EGF-LIKE 13.
FT DOMAIN 802 838 EGF-LIKE 14.
FT DOMAIN 840 902 EGF-LIKE 15.
FT DOMAIN 904 940 EGF-LIKE 16.
FT DOMAIN 942 978 EGF-LIKE 17.
FT DOMAIN 980 1021 EGF-LIKE 18.
FT DOMAIN 1021 1243 EGF-LIKE 19.
FT DOMAIN 1207 1243 EGF-LIKE 20.
FT DOMAIN 1481 1517 EGF-LIKE 21.
FT DOMAIN 1759 1795 EGF-LIKE 22.
FT DOMAIN 1797 1833 EGF-LIKE 23.
FT DOMAIN 1835 1871 EGF-LIKE 24.
FT DOMAIN 1874 1915 EGF-LIKE 25.
FT DOMAIN 1915 1951 EGF-LIKE 26.
FT DOMAIN 1953 1989 EGF-LIKE 27.
FT DOMAIN 1991 2029 EGF-LIKE 28.
FT DOMAIN 2030 2070 EGF-LIKE 29.
FT DISULFID 271 282 BY SIMILARITY.
FT DISULFID 276 291 BY SIMILARITY.
FT DISULFID 293 302 BY SIMILARITY.
FT DISULFID 310 321 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 333 342 BY SIMILARITY.
FT DISULFID 352 363 BY SIMILARITY.
FT DISULFID 357 374 BY SIMILARITY.
FT DISULFID 376 385 BY SIMILARITY.
FT DISULFID 392 403 BY SIMILARITY.
FT DISULFID 397 412 BY SIMILARITY.
FT DISULFID 414 424 BY SIMILARITY.
FT DISULFID 431 442 BY SIMILARITY.
FT DISULFID 436 451 BY SIMILARITY.
FT DISULFID 453 462 BY SIMILARITY.
FT DISULFID 468 479 BY SIMILARITY.
FT DISULFID 473 488 BY SIMILARITY.
FT DISULFID 490 499 BY SIMILARITY.
FT DISULFID 505 515 BY SIMILARITY.
FT DISULFID 509 520 BY SIMILARITY.
FT DISULFID 522 531 BY SIMILARITY.
FT DISULFID 549 562 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 571 580 BY SIMILARITY.
FT DISULFID 586 597 BY SIMILARITY.
FT DISULFID 591 602 BY SIMILARITY.
FT DISULFID 604 610 BY SIMILARITY.
FT DISULFID 613 624 BY SIMILARITY.
FT DISULFID 618 634 BY SIMILARITY.
FT DISULFID 636 645 BY SIMILARITY.
FT DISULFID 652 664 BY SIMILARITY.
FT DISULFID 659 673 BY SIMILARITY.
FT DISULFID 675 684 BY SIMILARITY.
FT DISULFID 691 702 BY SIMILARITY.
FT DISULFID 696 711 BY SIMILARITY.
FT DISULFID 713 722 BY SIMILARITY.
FT DISULFID 729 740 BY SIMILARITY.
FT DISULFID 734 749 BY SIMILARITY.
FT DISULFID 751 760 BY SIMILARITY.
FT DISULFID 767 778 BY SIMILARITY.
FT DISULFID 772 787 BY SIMILARITY.
FT DISULFID 789 799 BY SIMILARITY.
FT DISULFID 806 817 BY SIMILARITY.
FT DISULFID 811 826 BY SIMILARITY.
FT DISULFID 828 837 BY SIMILARITY.
FT DISULFID 844 855 BY SIMILARITY.
FT DISULFID 849 890 BY SIMILARITY.
FT DISULFID 892 901 BY SIMILARITY.
FT DISULFID 908 919 BY SIMILARITY.
FT DISULFID 913 928 BY SIMILARITY.
FT DISULFID 930 939 BY SIMILARITY.
FT DISULFID 946 957 BY SIMILARITY.
FT DISULFID 952 966 BY SIMILARITY.
FT DISULFID 968 977 BY SIMILARITY.
FT DISULFID 984 995 BY SIMILARITY.
FT DISULFID 989 1009 BY SIMILARITY.
FT DISULFID 1011 1020 BY SIMILARITY.
FT DISULFID 1211 1222 BY SIMILARITY.
FT DISULFID 1216 1231 BY SIMILARITY.
FT DISULFID 1233 1242 BY SIMILARITY.
FT DISULFID 1485 1496 BY SIMILARITY.
FT DISULFID 1490 1505 BY SIMILARITY.
FT DISULFID 1507 1516 BY SIMILARITY.
FT DISULFID 1763 1774 BY SIMILARITY.
FT DISULFID 1768 1783 BY SIMILARITY.
FT DISULFID 1785 1794 BY SIMILARITY.
FT DISULFID 1801 1812 BY SIMILARITY.
FT DISULFID 1806 1821 BY SIMILARITY.
FT DISULFID 1823 1832 BY SIMILARITY.
FT DISULFID 1839 1850 BY SIMILARITY.
FT DISULFID 1844 1859 BY SIMILARITY.
FT DISULFID 1861 1870 BY SIMILARITY.
FT DISULFID 1878 1889 BY SIMILARITY.
FT DISULFID 1883 1903 BY SIMILARITY.
FT DISULFID 1905 1914 BY SIMILARITY.
FT DISULFID 1919 1930 BY SIMILARITY.
FT DISULFID 1924 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1957 1968 BY SIMILARITY.
FT DISULFID 1962 1977 BY SIMILARITY.
FT DISULFID 1979 1988 BY SIMILARITY.
FT DISULFID 1995 2008 BY SIMILARITY.
FT DISULFID 2002 2017 BY SIMILARITY.
FT DISULFID 2019 2028 BY SIMILARITY.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```

Query Match          11.48;      Score 76; DB 1; Length 2139;
Best Local Similarity 30.98; Pred.No 8.1;
Matches 21; Conservative 7; Mismatches 28; Indels 12; Gaps 3;

QY   26 SPTKYNLLKESCI RNDCETGC-CORAPDNCSEHCAEKSGSLCQTQVFFGQVRACP 84
    II : : : : : I I : : : : :
Db   308 SPKACNPCEGSCLENSGNYQCFCDD-PNHSGQHCTEVNIHPICQTN-----P 356

QY   85 CLRNLTCTI 92
    II I I : :
Db   357 CLNNGACV 364

RESULT 13
NTCL_MOUSE
ID NTCL_MOUSE STANDARD; PRT; 2531 AA.
AC Q01705;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).
DE NOTCH1 GN MOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=EMBRYO;
RC MEDLINE-93194170; PubMed-8449489;
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RA Copeland N.G., Gridley T.;
RT "cloning, analysis, and chromosomal localization of Notch-1, a mouse
RL homologue of Drosophila Notch.";
RN Genomics 15:259-264(1993).
[2]
RN SEQUENCE OF 1551-2170 FROM N.A.
RP TISSUE=EMBRYO;
RC MEDLINE-93048835; PubMed-1425352;
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA Greenspan R.J., McMahon A.P., Gridley T.;
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
RT suggests an important role in early postimplantation mouse
development.";
RN Development 115:737-744(1992).
RL
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z11886; CAA77941.1; -.

```



```

FT CARBOHYD 260 260 PARTIAL.
FT VARSPLIC 229 301 MISSING (IN SHORT ISOFORM).
FT VARIANT 347 347 MISSING (IN CLONE HDLKAAG).
FT FTIG-VAR_002274.
FT CONFLICT 46 47 QP -> HV (IN REF. 1).
FT CONFLICT 108 108 G -> D (IN REF. 3).
FT SEQUENCE 383 AA; 41143 MW; 2C55C6E35B4415B CRC64;

Query Match 11.3%; Score 75; DB 1; Length 383;
Best Local Similarity 34.4%; Pred.No.2.4;
Matches 21; Conservative 6; Mismatches 22; Indels 12; Gaps 5;

QY 35 LKESCIRNODCETGCCORAPDNCESHCAEKGSGSLCOTQVFFGQYRAC---PCLRNLTG 91
DB 53 LDCQCVTSFGLGLGCE-PGQC--ICTD-GWDELCDRDV-----RACSSAPCANNGTC 103
QY 92 I 92
DB 104 V 104

RESULT 15
DISJ_BOTJA STANDARD; PRT; 571 AA.
AC P30431;
AD 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE PUTATIVE VENOM METALLOPROTEINASE JARARHAGIN PRECURSOR (EC 3.4.24.73)
DE (HF2-PROTEINASE) [CONTAINS: DISINTEGRIN] (FRAGMENT).
DE OS Bothrops jararaca (Jararaca).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-VENOM GLAND;
RX MEDLINE=93054601; PubMed=1395408;
RA Paine M.J.I., Desmond H.P., Theakston R.D.G., Crampton J.M.;
RT "Purification, cloning, and molecular characterization of a high
RT molecular weight hemorrhagic metalloprotease, jararhagin, from
RT Bothrops jararaca venom. Insights into the disintegrin gene family.";
RL J. Biol. Chem. 267:22869-22876(1992).
CC -I- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC ACTS IN HEMORRHAGE (BY SIMILARITY).
CC -I- FUNCTION: INHIBITS FIBRINOGEN INTERACTION WITH PLATELET RECEPTORS
CC EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. ACT BY BINDING TO THE
CC GLYCOPROTEIN IIB-IIIA RECEPTOR ON THE PLATELET SURFACE AND INHIBIT
CC AGGREGATION INDUCED BY ADP, THROMBIN, PLATELET-ACTIVATING FACTOR
CC AND COLLAGEN (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: CLEAVAGE OF 10-HIS-|-LEU-11, 14-ALA-|-LEU-15,
CC 16-TYR-|-LEU-17 AND 24-PHE-|-PHE-25 BONDS IN INSULIN B CHAIN.
CC -I- COFACTOR: BINDS ONE ZINC ION.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC -I- SIMILARITY: IN THE CENTRAL REGION; BELONGS TO THE DISINTEGRIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC of the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X68251; CAA48323.1; -.
CC DR PIR; S24789; S24789.
CC DR PIR; A44463; A44463.
CC DR HSP; P17494; 1KST.
CC DR MEROPS; M12.138; -.
CC DR INTERPRO; IPR000130; -.

```

```

DR INTERPRO; IPR001590; -
DR INTERPRO; IPR001762; -
DR INTERPRO; IPR002870; -
DR PFAM; PF01562; Pep_M12B_propep; 1.
DR PFAM; PF01421; Reptolysin; 1.
DR PFAM; PF00200; disintegrin; 1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
DR PROSITE; PS00214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Blood coagulation; Platelet;
KW Cell adhesion; Venom; Zymogen.
FT NON_TER 1 1
FT PROPEP <1 151 POTENTIAL.
FT CHAIN 152 355 JARAHAGIN.
FT CHAIN 356 447 DISINTEGRIN.
FT PROPEP 448 571
FT MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
FT METAL 295 295 ZINC (CATALYTIC) (PROBABLE).
FT ACT_SITE 296 296 BY SIMILARITY.
FT METAL 299 299 ZINC (CATALYTIC) (PROBABLE).
FT METAL 305 305 ZINC (CATALYTIC) (PROBABLE).
FT DISULFID 270 350 BY SIMILARITY.
FT DISULFID 310 317 BY SIMILARITY.
FT DISULFID 377 395 BY SIMILARITY.
FT DISULFID 379 390 BY SIMILARITY.
FT DISULFID 389 412 BY SIMILARITY.
FT DISULFID 403 409 BY SIMILARITY.
FT DISULFID 408 434 BY SIMILARITY.
FT DISULFID 421 441 BY SIMILARITY.
FT CARBOHYD 333 333 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 571 AA; 63982 MW; F5E549F8BF61177B CRC64;

```

```

Query Match 11.2%; Score 74.5; DB 1; Length 571;
Best Local Similarity 26.5%; Pred. No. 3.8;
Matches 30; Conservative 7; Mismatches 29; Indels 47; Gaps 7;

```

```

QY 32 LLELKESCI--RQDCETGCCQRAP-----DNCE----- 58
Db 370 LLEVGECDGTPENCQNECCDAATCKLKSQGGHGDCCCEQCKFSKSGTECRASMSECD 429
QY 59 --SHCAEKSGESLCQTVFF--GQYRACPLRLNLTCTIYSKNEKWLSTIAYGRC 107
Db 430 PAEHTGQSSE---CPADVEFKNGQ----PCLDNYGYCYNGN---CPIMVHQC 472

```

Search completed: March 24, 2001, 13:24:24
Job time: 170 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:22:26 ; Search time 47.63 Seconds
(without alignments)
86.866 Million cell updates/sec

Title: US-09-215-435-179

Perfect score: 666

Sequence: 1 MMLPQWLLLLLFLFFFL.....IAYGRCKIGROKLAKMFF 121

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36.*
1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	666	100.0	121	Y35930	Extended human sec
2	666	100.0	122	Y91513	Human secreted pro
3	619	92.9	120	Y12687	Human 5' EST secre
4	85.5	12.8	271	Y44934	Human dickkopf-1 h
5	82.5	12.4	179	Y73019	Human cysteine-ric
6	82.5	12.4	207	Y92074	Human DKK-2 splice
7	82.5	12.4	259	Y99360	Human PRO1316 (UNQ
8	82.5	12.4	259	Y92072	Murine DKK-2. Mus
9	82.5	12.4	259	Y92073	Human DKK-2. Homo
10	77	11.6	125	R25449	PDGF analogue #9.
11	77	11.6	125	R87521	Glioma cell PDGF A
12	76	11.4	802	R44242	F-spondin of chick

13	75	11.3	383	15	R56166	Neuroendocrine tum
14	75	11.3	1148	20	W87895	Human JAGGED2 prot
15	74	11.1	385	15	R56167	Neuroendocrine tum
16	74	11.1	807	20	Y04262	Human vascular smo
17	73	11.0	125	14	R44590	Monomeric PDGF-A.
18	73	11.0	807	14	R44241	F-spondin (FP5-9).
19	73	11.0	807	20	Y04263	Bovine vascular sm
20	73	11.0	807	20	Y04264	Rat vascular smoot
21	72.5	10.9	266	19	W73018	Human cysteine-ric
22	72.5	10.9	266	20	Y41757	Human PRO1008 prot
23	72.5	10.9	266	21	Y92071	Human DKK-1. Homo
24	72	10.8	125	14	R44584	Monomeric PDGF-A.
25	72	10.8	126	12	R12880	rPDGF A1 from pCFM
26	72	10.8	126	13	R25431	Glial PDGF A cha
27	72	10.8	211	12	R15645	PDGF D-1 A-chain p
28	72	10.8	211	14	R38918	Human Platelet Der
29	72	10.8	211	16	R44759	Human platelet-der
30	70.5	10.6	316	21	Y44862	Human taste cell s
31	70.5	10.6	1055	19	W44298	Human serrate 2 pr
32	70.5	10.6	1212	19	W44299	Human serrate 2.
33	70.5	10.6	1257	17	W05834	Human Serrate-2 (H
34	70.5	10.6	1257	21	Y59598	Human Serrate prot
35	69.5	10.4	224	19	W73017	Human cysteine-ric
36	69.5	10.4	224	21	Y92075	Human DKK-4. Homo
37	69.5	10.4	568	14	R44243	Xenopus F-spondin.
38	69.5	10.4	737	11	R05270	Amino acid sequenc
39	69	10.4	299	21	Y76014	Murine EGF family
40	69	10.4	299	21	Y76077	Murine EGF family
41	69	10.4	613	19	W73009	Cobra venom protea
42	69	10.4	621	19	W73013	Cobra venom moca
43	69	10.4	921	21	Y32297	Corn polycomb prot
44	68.5	10.3	79	17	R91704	NamNAP. Necator a
45	68.5	10.3	79	20	Y30412	Nematode extracted

ALIGNMENTS

RESULT 1

Y35930

ID Y35930 standard; Protein; 121 AA.

XX AC Y35930;

DT 13-SEP-1999 (first entry)

XX DE Extended human secreted protein sequence, SEQ ID NO. 179.

XX KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW genetic disease.

XX OS Homo sapiens.

XX PN WO9931236-A2.

XX PD 24-JUN-1999.

XX PF 17-DEC-1998; 98WO-IB02122.

XX PR 10-AUG-1998; 98US-0096116.

XX PR 17-DEC-1997; 97US-0069957.

XX PR 09-FEB-1998; 98US-0074121.

XX PR 13-APR-1998; 98US-0081563.

XX PA (GEST) GENSET.

XX XX

XX FI Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX WPI; 1999-385906/32.

XX DR N-PSDB; X97614.

XX New isolated human secreted proteins
XX PS Claim 9; Page 212; 516pp; English.
XX This sequence is encoded by an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.
XX SQ Sequence 121 AA;
Query Match 100.0%; Score 666; DB 20; Length 121;
Best Local Similarity 100.0%; Pred. No. 7.7e-58;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMLPQWLLLLLFLFFLLTRGSLSPTKYNLLEKESCIRNQDCETGCCORAPDNCEH 60
Db 1 mmlpqwlllllflfflltrgslsptkynlllelkesclrnqdcetgccqrapdnceh 60
QY 61 CAEKGSGSLCQTQVFFGQYRACPCRLNLTCTIYSKNEKWLSTIAYGRCQKIGROKLAKKMF 120
Db 61 caekgsgslcqtqvffgqyracpcrlnltctiysknekwlsaiygrcqkigrqlakkmf 120
QY 121 F 121
Db 121 f 121
RESULT 2
Y1513
ID Y1513 standard; Protein: 122 AA.
XX Y1513;
XX 29-JUN-2000 (first entry)
XX Human secreted protein sequence encoded by gene 63 SEQ ID NO:186.
XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiHIV; antiinflammatory; neutropic; neuroprotective; antiallergic;
KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW immune disease; inflammation; blood disorder; tumour.
XX Homo sapiens.
XX WO200006698-A1.
XX 10-FEB-2000.
XX 29-JUL-1999; 99WO-US17130.
XX 30-JUL-1998; 98US-0094657.
PR 05-AUG-1998; 98US-0095486.
PR 06-AUG-1998; 98US-0095454.
PR 06-AUG-1998; 98US-0095455.
PR 12-AUG-1998; 98US-0096319.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
WPI; 2000-195282/17.
N-PSDB; A26408.
New isolated human genes and the secreted polypeptides they encode,
useful for diagnosis and treatment of e.g. cancers, neurological
disorders, immune diseases, inflammation or blood disorders -
Claim 11; Page 499; 634pp; English.
The polynucleotide sequences given in A26346 to A26458 encode the human
secreted proteins given in Y91451 to Y91691. The human secreted proteins
can have activities based on the tissues and cells they are expressed in.
Examples of the activities are: cytostatic; immunosuppressive; antiHIV;
antiinflammatory; neutropic; neuroprotective; antiallergic; osteopathic;
antiarthritic; antibacterial; antidiabetic; antiasthma; antipsoriatic;
and cardiant. The polynucleotides and their corresponding secreted
proteins are useful for preventing, treating or ameliorating medical
conditions, e.g. by protein or gene therapy. Also pathological conditions
can be diagnosed by determining the amount of the proteins in a sample or
by determining the presence of mutations in the polynucleotides. Specific
uses are described for each of the polynucleotides, based on which
tissues they are most highly expressed in, and include developing
products for the diagnosis or treatment of cancer, tumours, and foetal
neurodegenerative disorders, developmental abnormalities and foetal
deficiencies, blood disorders, diseases of the immune system,
autoimmune diseases, hepatic and renal disease, inflammation,
allergies, Alzheimer's and behavioural disorders, schizophrenia,
osteoporosis, arthritis, infections, AIDS, sepsis, acne, psoriasis,
transplant rejection, diabetes, asthma, spinal cord injuries,
cardiovascular disorders, reproductive disorders, gastrointestinal
disorders, respiratory disorders and metabolic disorders. The proteins
or polynucleotides can also be used as food additives or preservatives.
The proteins are also useful for identifying their binding partners.
A26337 to A26345 and Y91450 are sequences used in the exemplification of
the present invention.
XX SQ Sequence 122 AA;
Query Match 100.0%; Score 666; DB 21; Length 122;
Best Local Similarity 100.0%; Pred. No. 7.8e-58;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMLPQWLLLLLFLFFLLTRGSLSPTKYNLLEKESCIRNQDCETGCCORAPDNCEH 60
Db 1 mmlpqwlllllflfflltrgslsptkynlllelkesclrnqdcetgccqrapdnceh 60
QY 61 CAEKGSGSLCQTQVFFGQYRACPCRLNLTCTIYSKNEKWLSTIAYGRCQKIGROKLAKKMF 120
Db 61 caekgsgslcqtqvffgqyracpcrlnltctiysknekwlsaiygrcqkigrqlakkmf 120
QY 121 F 121
Db 121 f 121
RESULT 3
Y12687
ID Y12687 standard; Protein: 120 AA.
XX Y12687;
XX 21-JUN-1999 (first entry)
XX Human 5' EST secreted protein SEQ ID NO:277.
XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

KW thrombolytic; anti-inflammatory; tumour inhibition.

XX OS Homo sapiens.

XX PN WO9906549-A2.

XX PD 11-FEB-1999.

XX PF 31-JUL-1998; 98WO-1B01231.

XX PR 01-AUG-1997; 97US-0905279.

XX PX (GEST) GENSET.

XX PA Duclert A, Dumas Milne Edwards J, Lacroix B;

XX PI WPI: 1999-153779/13.

XX DR N-PSDB; X51465.

XX New nucleic acids encoding human secreted proteins - obtained from
cDNA libraries derived from testis, ovary, uterus and spleen tissue

XX Claim 34; Page 387; 522pp; English.

XX X51459 to X51691 represent 5' expressed sequence tags (ESTs) for human
secreted proteins, and encode the proteins given in Y12681 to Y12913,
respectively. The proteins given represent the signal peptide and an
N-terminal fragment of a secreted protein. The nucleic acid sequences
can be used for producing secreted human gene products. The proteins
be used to develop products for diagnosis and therapy. They can also
obtained may have cytokine activity, cell proliferation/differentiation
activity, haematopoiesis regulating activity, tissue growth regulating
activity, reproductive hormone regulating activity, chemotactic/
chemokinetic activity, haemostatic and thrombolytic activity, receptor/
ligand activity, anti-inflammatory activity, tumour inhibition activity
or other activities. The products can be used in forensic, gene therapy
and chromosome mapping procedures. The sequences can also be used for
obtaining corresponding promoter sequences. The nucleic acids encoding
the signal peptide can be used for directing extracellular secretion of
a polypeptide or the insertion of a polypeptide into a membrane, or
importing a polypeptide into a cell.

XX Sequence 120 AA;

Query Match 92.9%; Score 619; DB 20; Length 120;

Best Local Similarity 95.8%; Pred. No. 2.8e-53;

Matches 113; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MMLPQWLLLLFFLLTRGSLSPKYNLLEKESCIRNQCETGCCQCAPDNCESH 60

Db 1 mmlpqwlllllfflltrgslspkynlllelxxxxngnqdcetgcccrapdncesh 60

QY 61 CAEKGSGSLCQTVFGQYRACPLRNLTCTIYSKNEKWLSTIAYGRQKGTGRKLAKK 118

Db 61 caekgsgslcqtqvffggyracplrnltctiysknekwlsiaaygrcqkigrqklark 118

RESULT 4

Y44934

ID Y44934 standard; Protein; 271 AA.

XX AC Y44934;

XX DT 23-MAY-2000 (first entry)

XX Human dickkopf-1 homolog 3 protein.

XX Human: dickkopf-1 homolog 3; DDKh-3; Wnt antagonist; probe; primer;

XX detect; quantitate; screening assay; immunogen; antibody.

XX Homo sapiens.

XX

FH Key Location/Qualifiers

FT Region 208..271

XX /note= "This region is claimed as SEQ ID NO: 2"

XX PN WO200006714-A1.

XX PD 10-FEB-2000.

XX PF 28-JUL-1999; 99WO-US16963.

XX PR 31-JUL-1998; 98US-0094947.

XX PR 31-JUL-1998; 98US-0094976.

XX PR 31-JUL-1998; 98US-0095217.

XX PA (ELIL) LILLY & CO ELI.

XX PI Edmonds BT;

XX WPI: 2000-195288/17.

XX Novel isolated Dickkopf-1 homolog 3 nucleic acid useful as a probe or
primer for detecting, quantitating and isolating gene transcripts in
screening assays -

XX Claim 6; Page 70; 72pp; English.

XX The present sequence is a human dickkopf-1 homolog 3 (DDKh-3) protein
which is selectively expressed in endocrine, and gastrointestinal tissues.
CC musculoskeletal, nervous, reproductive and gastrointestinal tissues.
CC It acts as Wnt antagonist. Dickkopf-1 has potent head inducing activity
in Xenopus. The DDKh-3 protein is used as immunogens in preparing
CC antibodies. The corresponding polynucleotide is used as a probe or primer
CC for detecting, quantitating or isolating gene transcripts e.g. in
CC screening assays for detecting allelic variants, or mutations. It is also
CC used for sense or antisense suppression of one or more DDKh-3 genes and
CC for identifying compounds that modulate transcription and translation
CC of the DDKh-3 genes.

XX Sequence 271 AA;

Query Match 12.8%; Score 85.5; DB 21; Length 271;

Best Local Similarity 26.1%; Pred. No. 0.46;

Matches 23; Conservative 9; Mismatches 29; Indels 27; Gaps 5;

QY 39 CIRNQDCECTGCCQCAPDNCESH-----CAEKGSGSLCQTVFGQ-----YRACPLRN 88

Db 195 clrssdcaaglc-----carhfwtkickpvlhgdgevtckqrkkghgleifgrcdcaqg 248

QY 89 LTCIYSKNEKWLSTIAYGR-----CQKI 110

Db 249 lsc-----kvwkdatysskarllhvcqki 271

RESULT 5

W73019

ID W73019 standard; Protein; 179 AA.

XX AC W73019;

XX DT 18-JAN-1999 (first entry)

XX Human cysteine-rich secreted protein CRSP-4.

XX CRSP-4; cysteine-rich secreted protein 4; tumour; cancer; leukaemia;

XX tissue repair; wound healing; infection; Parkinson's disease;

XX Alzheimer's disease; Huntington's chorea; multiple sclerosis;

XX amyotrophic lateral sclerosis; pontine myelinolysis;

XX human immunodeficiency associated myelopathy; bulbar palsy;

XX spinal muscular atrophy; primary lateral sclerosis; poliomyelitis;

XX Razio-Londe syndrome; Charcot-Marie-Tooth disease; therapy;

XX diagnosis; drug screening; human; signal transduction;

XX cell differentiation; cell proliferation.

XX Homo sapiens.
 XX Key Location/Qualifiers
 XX Domain 1..47
 XX Peptide /note= "cysteine-rich domain"
 XX 48..95
 XX Domain /note= "spacer"
 XX 96..176
 XX /note= "cysteine-rich domain"
 XX WO9846755-A1.
 XX 22-OCT-1998.
 XX 16-APR-1998; 98WO-US07894.
 XX 20-JAN-1998; 98US-0009802.
 XX 16-APR-1997; 97US-0843704.
 XX 17-APR-1997; 97US-0842898.
 XX 15-JAN-1998; 98US-0071589.
 XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 XX McCarthy SA;
 XX WPI: 1998-568730/48.
 XX N-PSDB; V07909.
 XX New isolated cysteine-rich secreted proteins - used to develop
 XX products for treating, e.g. hyperproliferative disorders, cancers,
 XX wounds, infectious lesions, degenerative lesions or demyelinating
 XX diseases
 XX Claim 8; Page 105; 142pp; English.
 XX This is the amino acid sequence of novel human cysteine-rich
 XX secreted protein 4 (CRSP-4), as deduced from a partial cDNA
 XX sequence (see V07909). CRSP-4 mRNA is expressed in several adult
 XX tissues, with highest expression in heart, brain, lung, placenta
 XX and skeletal muscle. CRSP-1, -2, -3 and -4 (see W73016-19) are
 XX claimed, characterised by the presence of cysteine-rich domains
 XX joined by spacer regions. CRSPs have at least 1 of the following
 XX activities: (i) modulation of cellular signal transduction, either
 XX in vitro or in vivo (e.g. antagonism of the activity of members of
 XX the wnt family of secreted proteins or suppression of wnt-dependent
 XX signal transduction); (ii) regulation of communication between cells
 XX (e.g. regulation of wnt-dependent cell-cell interactions); (iii)
 XX regulation of expression of genes whose expression is modulated by
 XX binding of CRSP (e.g. CRSP-1) to a receptor; (iv) regulation of gene
 XX transcription in a cell involved in development or differentiation;
 XX either in vitro or in vivo (e.g. induction of cell differentiation);
 XX (v) regulation of cellular proliferation, either in vitro or in vivo
 XX (e.g. induction of cellular proliferation or inhibition of
 XX proliferation as in the case of suppression of tumorigenesis (e.g.
 XX suppression of glioblastoma formation)); (vi) formation and
 XX maintenance of ordered spatial arrangements of differentiated
 XX tissues in vertebrates, both adult and embryonic (e.g. induction of
 XX head formation during vertebrate development or maintenance of
 XX haematopoietic progenitor cells); (vii) modulation of cell death,
 XX such as stimulation of cell survival; (ix) regulating cell
 XX migration, and/or (xii) immune modulation. Products of the
 XX invention can be used for e.g. treating hyperproliferative diseases
 XX such as neoplastic and hyperplastic disease, e.g. various cancers
 XX and leukaemias, and fibroproliferative disorders, malignancy, or
 XX premalignant and benign conditions, to stimulate tissue repair or
 XX wound healing or treat spinal cord injuries, brain injuries, lesions
 XX associated with surgery, ischemic lesions, malignant lesions,
 XX infectious lesions, degenerative lesions (Parkinson's disease,
 XX Alzheimer's disease, Huntington's chorea, amyotrophic lateral
 XX sclerosis), demyelinating diseases (multiple sclerosis, human
 XX immunodeficiency associated myelopathy, pontine myelinolysis),
 XX motor neuron injuries, progressive spinal muscular atrophy,

CC progressive bulbar palsy, primary lateral sclerosis, infantile and
 CC juvenile muscular atrophy, progressive bulbar paralysis of childhood
 CC (Fazio-Londe syndrome), poliomyelitis, and hereditary motor-sensory
 CC neuropathy (Charcot-Marie-Tooth disease). The products can also be
 CC used for detection, diagnosis and drug screening.
 XX Sequence 179 AA;
 SQ
 Query Match 12.4%; Score 82.5; DB 19; Length 179;
 Best Local Similarity 25.6%; Pred. NO. 0.59;
 Matches 23; Conservative 10; Mismatches 30; Indels 27; Gaps 5;
 QY 37 ESCIRNQDCTCCQAPDNCESH-----CAEKSEGLCQTQVFFGQ-----YRACPL 86
 Db 101 dpcirssdciegc-----carhfwtkickpvlhgvevctkqrkkgshgileifrcdca 154
 QY 87 RNLTCTYSKNEKWLSTAYGR-----COKI 110
 Db 155 kglsc-----kvkdatysskarlhvqcqi 179
 RESULT 6
 Y92074
 ID Y92074 standard; Protein: 207 AA.
 XX AC Y92074;
 XX DT 01-AUG-2000 (first entry)
 XX DE Human DKR-2 splice variant, DKR-2a.
 XX DKR-2a; dkk-1; dickkopf-1; antagonist; wnt-8 signaling; morphogenesis;
 XX growth factor; cytoskeletal; sonic hedgehog; tissue differentiation.
 XX Homo sapiens.
 OS WO200018914-A2.
 PN 06-APR-2000.
 PD 17-SEP-1999; 99WO-US21647.
 PF 25-SEP-1998; 98US-0161241.
 PR (AMGE-) AMGEN INC.
 PA Bass MB, Sullivan JK, Theill LE, Wang D;
 PI WPI: 2000-293153/25.
 DR N-PSDB; A08843.
 XX New nucleic acid molecule encoding a biologically active DKR
 XX polypeptide, useful in treatment of cancer, e.g. mammary tumors and
 XX stem cell tumors
 XX Claim 18; Page 130-131; 143pp; English.
 XX Y92069-75 are novel mouse and human DKR polypeptides.
 XX The human DKR-2a gene is a splice variant of DKR-2 and seems to be
 XX missing an exon. Human DKR-2 has about 95 percent homology to murine
 XX DKR-2. DKR-1 is a human ortholog of dkk-1 (dickkopf-1), a novel gene
 XX identified in Xenopus and mouse, purportedly an antagonist of wnt-8
 XX signaling. DKR-2, -3 and -4 are each related to DKR-1 by their cysteine
 XX pattern. Dkk-1 is also involved in morphogenesis in the developing
 XX embryo, and therefore a growth factor, by inference DKR polypeptides
 XX are also growth factors. The DKR polypeptides are useful for treating
 XX cancer, e.g. mammary tumors, stem cell tumors, or other cancers in which
 XX the wnt and/or sonic hedgehog (shh) signal transduction pathways are
 XX activated. They can also be used to enhance tissue differentiation, such
 XX as bone formation and hematopoietic cell formation.
 XX Sequence 207 AA;
 SQ

PR 17-NOV-1998; 98US-0108802.
 PR 17-NOV-1998; 98US-0108806.
 PR 17-NOV-1998; 98US-0108807.
 PR 17-NOV-1998; 98US-0108867.
 PR 17-NOV-1998; 98US-0108925.
 PR 18-NOV-1998; 98US-0108848.
 PR 18-NOV-1998; 98US-0108849.
 PR 18-NOV-1998; 98US-0108850.
 PR 18-NOV-1998; 98US-0108851.
 PR 18-NOV-1998; 98US-0108852.
 PR 18-NOV-1998; 98US-0108858.
 PR 18-NOV-1998; 98US-0108904.
 XX (GETH) GENENTECH INC.
 XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
 XX WPI: 2000-237871/20.
 DR N-PSDB; A37042.
 XX New mammalian DNA sequences encoding transmembrane, receptor or
 PT secreted PRO polypeptides useful for screening of potential peptide or
 PT small molecule inhibitors of the relevant receptor/ligand interactions
 XX Claim 12: Fig 42; 773pp; English.
 XX A37022 to A37144 encode the new isolated human transmembrane, receptor
 CC or secreted PRO polypeptides given in Y99340 to Y99462. The
 CC transmembrane and receptor PRO proteins can be used for screening of
 CC potential peptide or small molecule inhibitors of the relevant
 CC receptor/ligand interactions. The polypeptides and nucleotide sequences
 CC encoding them have various industrial applications, including uses as
 CC pharmaceutical and diagnostic agents. A37145 to A37330 represent
 CC PCR primers and hybridisation probes used in the isolation of the PRO
 CC polypeptides from the present invention.
 XX Sequence 259 AA;
 SQ
 Query Match 12.4%; Score 82.5; DB 21; Length 259;
 Best Local Similarity 25.6%; Pred. No. 0.87;
 Matches 23; Conservative 10; Mismatches 30; Indels 27; Gaps 5;
 QY 37 ESCIRNQCETGCCQAPDNCESH-----CAEKGSEGLCQTQVFFGQ-----YRACPCL 86
 Db 181 dpcrlrddciefc-----carhfwtkickpvlhggevtckqrkkghgleifqrdca 234
 QY 87 RNLTCTYSKNEKWLSTAYGR-----CQKI 110
 Db 235 kgisc-----kvwkdatyskskarlhvcqki 259
 RESULT 8
 Y92072
 ID Y92072 standard; Protein; 259 AA.
 AC Y92072;
 XX
 DT 01-AUG-2000 (first entry)
 DE Murine DKR-2.
 XX DKR-2; dkk-1; dickkopf-1; antagonist; wnt-8 signaling; morphogenesis;
 KW growth factor; cytotostatic; sonic hedgehog; tissue differentiation.
 XX Mus sp.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..33
 FT /label= signal_peptide
 FT Cleavage-site 33..34
 FT Protein 34..259
 FT /label= mature_protein

XX WO200018914-A2.
 XX 06-APR-2000.
 XX 17-SEP-1999; 99WO-US21647.
 XX 25-SEP-1998; 98US-0161241.
 XX (AMGE-) AMGEN INC.
 XX Bass MB, Sullivan JK, Theill LE, Wang D;
 XX WPI: 2000-293153/25.
 DR N-PSDB; A08841.
 XX New nucleic acid molecule encoding a biologically active DKR
 PT polypeptide, useful in treatment of cancer, e.g. mammary tumors and
 PT stem cell tumors
 XX Claim 18; Page 128-129; 143pp; English.
 XX Y92069-75 are novel mouse and human DKR polypeptides.
 CC The murine DKR-2 gene has about 38 percent homology to murine dkk-1
 CC (dickkopf-1), a novel gene identified in Xenopus and mouse, purportedly
 CC an antagonist of wnt-8 signaling.
 CC Dkk-1 is also involved in morphogenesis in the developing embryo, and
 CC therefore a growth factor, by inference DKR polypeptides are also
 CC growth factors. The DKR polypeptides are useful for treating cancer,
 CC e.g. mammary tumors, stem cell tumors, or other cancers in which the wnt
 CC and/or sonic hedgehog (shh) signal transduction pathways are activated.
 CC They can also be used to enhance tissue differentiation, such as bone
 CC formation and hematopoietic cell formation.
 XX Sequence 259 AA;
 SQ
 Query Match 12.4%; Score 82.5; DB 21; Length 259;
 Best Local Similarity 25.6%; Pred. No. 0.87;
 Matches 23; Conservative 10; Mismatches 30; Indels 27; Gaps 5;
 QY 37 ESCIRNQCETGCCQAPDNCESH-----CAEKGSEGLCQTQVFFGQ-----YRACPCL 86
 Db 181 dpcrlrddciefc-----carhfwtkickpvlhggevtckqrkkghgleifqrdca 234
 QY 87 RNLTCTYSKNEKWLSTAYGR-----CQKI 110
 Db 235 kgisc-----kvwkdatyskskarlhvcqki 259
 RESULT 9
 Y92073
 ID Y92073 standard; Protein; 259 AA.
 XX
 AC Y92073;
 XX
 DT 01-AUG-2000 (first entry)
 DE Human DKR-2.
 XX DKR-2; dkk-1; dickkopf-1; antagonist; wnt-8 signaling; morphogenesis;
 KW growth factor; cytotostatic; sonic hedgehog; tissue differentiation.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..33
 FT /label= signal_peptide
 FT Cleavage-site 33..34
 FT Protein 34..259
 FT /label= mature_protein
 XX WO200018914-A2.
 PN

XX PD 06-APR-2000.
 XX PF 17-SEP-1999; 99WO-US21647.
 XX PR 25-SEP-1998; 98US-0161241.
 XX PA (AMGE-) AMGEN INC.
 XX PI Bass MB, Sullivan JK, Theill LE, Wang D;
 XX PT WPI; 2000-293153/25.
 XX DR N-PSDB; A08842.
 XX PT New nucleic acid molecule encoding a biologically active DKR
 XX PT polypeptide, useful in treatment of cancer, e.g. mammary tumors and
 XX PT stem cell tumors
 XX PS Claim 18; Page 129-130; 143pp; English.
 XX CC Y92069-75 are novel mouse and human DKR polypeptides.
 CC The human DKR-2 gene has about 95 percent homology to murine DKR-2.
 CC DKR-1 is a human ortholog of dkk-1 (dickkopf-1), a novel gene identified
 CC in Xenopus and mouse, purportedly an antagonist of wnt-8 signaling.
 CC DKR-2, -3 and -4 are each related to DKR-1 by their cysteine pattern.
 CC Dkk-1 is also involved in morphogenesis in the developing embryo, and
 CC therefore a growth factor, by inference DKR polypeptides are also
 CC growth factors. The DKR polypeptides are useful for treating cancer,
 CC e.g. mammary tumors, stem cell tumors, or other cancers in which the wnt
 CC and/or sonic hedgehog (shh) signal transduction pathways are activated.
 CC They can also be used to enhance tissue differentiation, such as bone
 CC formation and hematopoietic cell formation.
 XX SQ Sequence 259 AA;

Query Match 12.4%; Score 82.5; DB 21; Length 259;
 Best Local Similarity 25.6%; Pred. No. 0.87;
 Matches 23; Conservative 10; Mismatches 30; Indels 27; Gaps 5;
 QY 37 ESCIRNDCETGCCORAPDNCEH-----CAEKGSEGLCQTQVFFGQ-----YRACPCL 86
 Db 181 dpcrlssdclegfc-----carhfwtkickpvlhgdvctkqrkkgshgleifqrcdca 234
 QY 87 RNLTCIYSKNEKWLSTAYGR-----CQKI 110
 Db 235 kgisc-----kvwkdatysskarlhvcqki 259

RESULT 10
 R25449
 ID R25449 standard; Protein; 125 AA.
 XX AC R25449;
 XX DT 13-JAN-1993 (first entry)
 XX DE PDGF analogue #9.
 XX KW platelet derived growth factor; chemotactic; mitogenic; fibroblasts;
 KW wound healing; dermal ulcers; lacerations; abrasions;
 KW surgical wounds; burns; defined culture media; v-sis protein; p28sis;
 KW neoplasm; cancer; tumour; inhibit atherosclerosis.
 XX OS Synthetic.
 XX PN US5128321-A.
 XX PD 07-JUL-1992.
 XX PF 08-AUG-1988; 88US-0230190.
 XX PR 12-OCT-1984; 84US-0660496.

PR 25-FEB-1985; 85US-0705175.
 PR 13-AUG-1986; 86US-0896485.
 PR 15-DEC-1986; 86US-0941970.
 XX (ZYMO) ZYMOGENETICS INC.
 XX PI Kelly JD, Murray MJ;
 XX DR WPI; 1992-249442/30.
 XX PT New platelet-derived growth factor analogues - comprising protein
 PT homo:dimers having two polypeptide chains with sequences the same
 PT as the A- and opt. B chain obtd. in large amts. used to enhance the
 PT healing of wounds
 XX PS Disclosure; Fig 3A; 25pp; English.
 XX CC This sequence represents a new platelet-derived growth factor
 CC analogue chain . The analogue functions as a homodimer of two chains,
 CC each of the chains being a mosaic of amino acid sequences identical
 CC to portions of the A- or B-chains of PDGF, the protein being
 CC chemotactic or mitogenic for fibroblasts. The homodimers have the
 CC same biological activity as PDGF, but can be obtained in large
 CC amts. by recombinant DNA techniques. These PDGF analogues can be
 CC used in a compsn. to enhance the healing of wounds eg. dermal
 CC ulcers, superficial wounds and lacerations, abrasions, surgical
 CC wounds and some burns. The PDGF analogues can also be used as
 CC components of defined, serum-free culture media. They can also be
 CC used in elucidating the putative role of the vsis protein p28sis
 CC in the neoplastic process or for developing inhibitors or
 CC designing specific therapeutic approaches which prevent or
 CC interfere with the in vivo activity of PDGF in individuals with
 CC atherosclerosis.
 XX SQ Sequence 125 AA;

Query Match 11.6%; Score 77; DB 13; Length 125;
 Best Local Similarity 22.8%; Pred. No. 1.4;
 Matches 26; Conservative 22; Mismatches 48; Indels 18; Gaps 5;
 QY 17 LFLTRGSLSPKYNNLELKESCIRNDCETGCCORAPDNCE-----SHCAEK-----G 65
 Db 16 iyeiprsqvdtsanfl-iwppcvevkrc-tgcqctssvkcpshrsvkvakveyvr 73
 QY 66 SEGSLCQTQVFFGQYRACPCLRNLTCTIYSKNEKWLSTAYGECQKIGROKLAKKM 119
 Db 74 kpkkiqevqrleehleacatt-----slndpyreedtgrpresgkkrkrkl 122

RESULT 11
 R87521
 ID R87521 standard; peptide; 125 AA.
 XX AC R87521;
 XX DT 21-JUN-1996 (first entry)
 XX DE Glioma cell PDGF A chain.
 XX KW Platelet-derived growth factor; PDGF; unglycosylated protein; burn;
 KW wound healing; non-healing dermal ulcer; superficial wound; laceration;
 KW abrasion; surgical wound; advanced age; diabetes; cancer; steroid;
 KW anti-inflammatory drug; anticoagulant; homodimer.
 XX OS Synthetic.
 XX PN US5474982-A.
 XX PD 12-DEC-1995.
 XX PF 13-AUG-1986; 86US-0896485.
 XX

PR 08-AUG-1988; 88US-0230190.
 PR 13-AUG-1986; 86US-0896485.
 PR 15-DEC-1986; 86US-0941970.
 PR 30-JUN-1992; 92US-0906544.
 PR 12-JUL-1994; 94US-0273779.
 XX (ZYMO) ZYMOGENETICS INC.
 XX Kelly JD, Murray MJ;
 XX WPI; 1996-039529/04.
 XX New platelet-derived growth factor analogues - useful for promoting
 PT wound healing
 XX Example 3; Fig 3b; 25pp; English.
 XX R87513-R87521 are representative peptides of the invention. These
 CC sequences are all mosaics of the platelet-derived growth factor (PDGF) A
 CC chain. This sequence represents a glioma cell A chain variant, which can
 CC be used as a peptide of the invention. These sequences are used to form
 CC the novel unglycosylated protein homodimers of the invention. These
 CC homodimers can be used in compositions for enhancing the wound healing
 CC process. The protein is expected to accelerate the healing process in a
 CC broad spectrum of wound conditions including disruptions of the dermal
 CC layer of the skin (such as non-healing dermal ulcers, superficial wounds
 CC and lacerations, abrasions, surgical wounds, and some burns). These
 CC sequences are especially used in enhancing the wound-healing process in
 CC conditions where the normal wound healing process is suppressed or
 CC inhibited, such as advanced age, diabetes, cancer, and treatment with
 CC anti-inflammatory drugs, steroids or anticoagulants.
 XX Sequence 125 AA;
 SQ

Query Match 11.6%; Score 77; DB 17; Length 125;
 Best Local Similarity 22.8%; Pred. No. 1.4;
 Matches 26; Conservative 22; Mismatches 48; Indels 18; Gaps 5;
 QY 17 LFLRLRGSLPTKYNLLEKESCIRNDCETGCCQAPDNCE----SHCAEK-----G 65
 Db 16 iyeiprsqvdtsanfl-lwppcvvkrcl-gccqtssvkqpsrvhrsvkvakveyvr 73
 QY 66 SEGSLCQTVFGQVRACPLRLNLTCTIYSKNEKWLSTIAYGRCQKIGROKLAKKM 119
 Db 74 kfpklkevqvrleehleacatt-----slpdyreedtgrpresgkkrkrkrl 122

RESULT 12
 R44242
 ID R44242 standard; Protein; 802 AA.
 XX
 AC R44242;
 XX
 DT 03-MAY-1994 (first entry)
 XX F-spondin of chicken.
 DE Thrombospondin; TSR; Thrombospondin type I repeat; F-spondin;
 KW neurodevelopment; nerve; axon; adhesion; outgrowth.
 KW Gallus gallus.
 XX WO9320196-A.
 PN 14-OCT-1993.
 XX 02-APR-1993; 93WO-US03164.
 XX 02-APR-1992; 92US-0862021.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX

PI Jessell TM, Klar A;
 XX WPI; 1993-336904/42.
 DR N-PSDB; Q52675.
 XX New vertebrate F-spondin protein - used for attaching nerve cells
 PT to a matrix, stimulating growth of nerve cells or regenerating
 PT nerve cells
 XX Claim 18; Page 68-70; 103pp; English.
 XX F-spondin is useful for adhesion and outgrowth of axons. It can be
 CC used for attaching nerve cells to a matrix, stimulating growth of
 CC nerve cells or regenerating nerve cells. F-spondin nucleic acid
 CC can be used to develop probes to study neurodevelopment.
 CC Antibodies to the F-spondin can be used for determining the
 CC localisation of the protein in the nervous system and in assessing
 CC its function.
 XX Sequence 802 AA;
 SQ
 Query Match 11.4%; Score 76; DB 14; Length 802;
 Best Local Similarity 21.8%; Pred. No. 12;
 Matches 22; Conservative 17; Mismatches 44; Indels 18; Gaps 4;
 QY 34 ELKESCIRNDCETGCCQAP----DNCESHCAE-----KGSEGLCQTVFGF 78
 Db 538 eetekclvneecspssclvtewgewdecscqctgmkrhrmikmtpadgsmckaettea 597
 QY 79 QYRACPLRLNLTCTIYSKNEKWLSTIAYGRCQKIGROKLAK 117
 Db 598 ekcmpechtipcllspwsewscsvtcgkgmrt-rgrmlk 637

RESULT 13
 R56166
 ID R56166 standard; Protein; 383 AA.
 XX
 AC R56166;
 XX 26-JAN-1995 (first entry)
 DT Neuroendocrine tumor dlk.
 DE Dlk; neuroendocrine tumor; monoclonal antibody.
 KW Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..23
 FT /label= Sig_peptide
 FT Modified-site 62
 FT /label= N-myristylation_site
 FT Modified-site 72
 FT /label= N-myristylation_site
 FT Modified-site 100
 FT /label= N-glycosylation_site
 FT Modified-site 101
 FT /label= N-myristylation_site
 FT Modified-site 109
 FT /label= N-myristylation_site
 FT Modified-site 120
 FT /label= protein-kinase-C-phosphorylation_site
 FT Modified-site 187
 FT /label= hydroxylation_site
 FT Modified-site 288
 FT /label= protein-kinase-C-phosphorylation_site
 FT Modified-site 302
 FT /label= N-myristylation_site
 FT Modified-site 312
 FT /label= N-myristylation_site
 XX

PN WO9413701-A.

XX 23-JUN-1994.

XX 10-DEC-1993; 93WO-US12015.

XX 11-DEC-1992; 92US-0989537.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX (USSH) US SEC DEPT. HEALTH.

XX Laborda J;

XX WPI; 1994-217806/28.

XX N-PSDB; Q68221.

XX A Dlk polypeptide encoded by dlk polynucleotide mol - can be used

PT in detecting the existence of primary or secondary neuro

PT endocrine tumour

XX Disclosure; Fig. 1; 24pp; English.

XX A cDNA clone was isolated from a phage lambda-gtl0 human adrenal

CC gland library using mouse dlk as hybridization probe. The human

CC (Q68221) and mouse (Q68222) dlk cDNAs showed 82.2% identity. The

CC encoded human (R56166) and mouse (R56167) Dlk proteins are used to

CC detect neuroendocrine tumors and to produce monoclonal antibodies

CC useful for diagnosis and therapy.

XX Sequence 383 AA;

SQ Query Match 11.3%; Score 75; DB 15; Length 383;

Best Local Similarity 34.4%; Pred. No. 6.9;

Matches 21; Conservative 6; Mismatches 22; Indels 12; Gaps 5;

QY 35 LKESCIRNQCETGCCORAPDNCESHCAERKSGSLCOTQVFFQYRAC---PCLRNLTG 91

Db 53 lcdqcvspgclhgcge-p9qc--ictd-gwdgelcdrdv-----racssapcanngtc 103

QY 92 I 92

Db 104 v 104

RESULT 14

W87895

ID W87895 standard; Protein; 1148 AA.

XX W87895;

XX 26-APR-1999 (first entry)

XX Human JAGGED2 protein.

XX JAGGED2; JAGGED2; human; notch ligand; syndactyly;

XX cleaft palate; cleft lip; diagnosis; stem cell;

XX progenitor cell; haematopoiesis; cell differentiation.

XX Homo sapiens.

OS WO9858958-A2.

XX 30-DEC-1998.

XX 25-JUN-1998; 98WO-US13207.

XX 25-JUN-1997; 97US-0882046.

XX (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.

PA (UNIW) UNIV WASHINGTON.

XX Hood L, Krantz ID, Li L, Spinner NB;

PI

XX

DR WPI; 1999-081220/07.

XX N-PSDB; V63754.

XX New Jagged peptides for inhibiting differentiation of progenitor

PT cells - also used for maintaining these cells in undifferentiated

PT state, e.g. for haematopoietic reconstitution

XX Claim 6; Fig 1B; 101pp; English.

XX This is the amino acid sequence of human JAGGED2 (hJAGGED2), a

XX membrane-bound ligand with a large extracellular domain and a

XX relative short intracellular domain. The polypeptide also has a

XX Delta/Serrate/Lag-2 domain, 15 EGF-like repeats and a

XX transmembrane domain characteristic of membrane-bound Notch

XX ligands. Genetic polymorphisms in the human JAGGED2 gene (see

XX V63754) are associated with syndactyly and cleft palate or lip.

XX A method of diagnosing a syndrome characterised by such conditions

XX is disclosed. The invention also relates to new human JAGGED2

XX protein (see W87894) and polynucleotides (see V63753) and their

XX used for inhibiting progenitor cell differentiation and diagnosing

XX Alagille syndrome.

XX Sequence 1148 AA;

SQ Query Match 11.3%; Score 75; DB 20; Length 1148;

Best Local Similarity 23.2%; Pred. No. 21;

Matches 23; Conservative 11; Mismatches 47; Indels 18; Gaps 4;

QY 31 NLELKESCIRNQCETGCCORAPDNCESHCAERKSGSLCOTQVFFQYRACPLRNLT 90

Db 293 nckeahactsnpcanggschevpsgfechc-psgwsqptcaldi--decasnpcaaggt 349

QY 91 CIYSKN-----EKWLSI-----AYRCQKIGROK 114

Db 350 cvdqvdfecicpeqwgatcqldvndcrgcqbhggtck 388

RESULT 15

R56167

ID R56167 standard; Protein; 385 AA.

XX R56167;

XX 26-JAN-1995 (first entry)

XX Neuroendocrine tumor dlk.

XX Dlk; neuroendocrine tumor; monoclonal antibody.

XX Mus sp.

XX Key

XX Peptide

XX Location/Qualifiers

XX 1..23

XX /label= Sig_peptide

XX Modified-site

XX 62

XX /label= N-myristylation_site

XX Modified-site

XX 100

XX /label= N-glycosylation_site

XX Modified-site

XX 109

XX /label= N-myristylation_site

XX Modified-site

XX 120

XX /label= protein-kinase-C_phosphorylation_site

XX Modified-site

XX 189

XX /label= hydroxylation_site

XX Modified-site

XX 269

XX /label= protein-kinase-C_phosphorylation_site

XX Modified-site

XX 290

XX /label= protein-kinase-C_phosphorylation_site

XX Modified-site

XX 295

XX /label= N-glycosylation_site

XX Modified-site

XX 304

	Query Match	11.18;	Score 74;	DB 15;	Length 385;
	Best Local Similarity	31.18;	Pred. No. 8.7;	Mismatches	Gaps
	Matches	19;	Conservative	Indels	23; Indels
QY	35	LKESCIRNQDCTGCQRAPDNCESHCAEKSEGLCQTQVFFGQYRAC----	PCLRNLTC	91	
Db	53	lckdckvtapgvngvc-kepwgc-----lckdgwdgxfceldv-----ractstpcanngtc	103		
QY	92	I	92		
Db	104	v	104		

Search completed: March 24, 2001, 13:22:28
Job time: 59 sec

[illegible]


```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA Monaghan P.A., Kioschis P., Wu W., Zuniga A., Bock D., Poustka A.,
RA Delius H., Niehrs C.;
RT "Dickkopf genes are co-ordinately expressed in mesodermal lineages.";
RL Mech. Dev. 87:45-56(1999).
DR EMBL: AJ243963; CAB60110.1; -.
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 259 DICKKOPF-2.
SQ SEQUENCE 259 AA; 28416 MW; EAAB76F2D2C9780D CRC64;

Query Match 12.4%; Score 82.5; DB 11; Length 259;
Best Local Similarity 25.6%; Pred. No. 0.064;
Matches 23; Conservative 10; Mismatches 30; Indels 27; Gaps 5;

QY 37 ESCIRNDCETGCCORAPDNCEH-----CAEKSGSLCQTQVFFGQ-----YRACPCL 86
DB 181 DPLRSSDCIDGFC-----CARHFWTKICKPVHOGVECTORKKGGHGLEIFQRCDCA 234

QY 87 RNLTCTYSKNEKWLSTAYGR-----CQKI 110
DB 235 KGLSC-----KVKWDATYSSKARLHVCQKI 259

RESULT 6
ID 070534 PRELIMINARY; PRT; 383 AA.
AC 070534;
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TEMBLrel. 14, Last annotation update)
DE ZOG.
GN ZOG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RA Haider S.K., Takemori H., Akira W., Yasuki N., Mitsuhiro O.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: D84336; BAA25881.1; -.
DR HSSP: P00740; 1IXA.
DR INTERPRO: IPR000152; -.
DR INTERPRO: IPR000561; -.
DR INTERPRO: IPR000742; -.
DR INTERPRO: IPR001881; -.
DR PFAM: PF00008; EGF_5.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_5.
DR PROSITE: PS01186; EGF_2; 6.
KW Glycoprotein.
SQ SEQUENCE 383 AA; 41149 MW; 123724C9801A9FD5 CRC64;

Query Match 12.3%; Score 82; DB 11; Length 383;
Best Local Similarity 31.1%; Pred. No. 0.11;
Matches 19; Conservative 9; Mismatches 21; Indels 12; Gaps 4;

QY 35 LKESCIRNDCETGCCORAPDNCEHSCAEGSGSLCQTQVFFGQYRAC---PCLRNLTG 91
DB 53 LCEKCVTSPGCVNGLCEE-PWQC---VCKEGWDGKFCFCEIDI-----RACTSTPCANNNGTC 103

QY 92 I 92
DB 104 V 104

RESULT 8
ID Q9YHU2 PRELIMINARY; PRT; 1254 AA.
AC Q9YHU2;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE SERRATEB.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi.
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RA Haddon C., Jiang Y.-J., Smithers L., Lewis J.;
RT "Delta-Notch signalling and the patterning of sensory cell

```

```

RESULT 7
Q62779 PRELIMINARY; PRT; 383 AA.
ID Q62779;
AC Q62779;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TEMBLrel. 14, Last annotation update)
DE PREADIPOCYTE FACTOR 1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-WISTAR; TISSUE-ISLETS OF LANGERHANS;
RX MEDLINE-97419009; PubMed-9275085;
RA Carlsson C., Tornehave D., Lindberg K., Galante P., Billestrup N.,
RA Michelsen B., Larsson L.I., Nielsen J.H.;
RT "Growth hormone and prolactin stimulate the expression of rat
RT preadipocyte factor-1/delta-like protein in pancreatic islets;
RT molecular cloning and expression pattern during development and growth
RT of the endocrine pancreas.";
RL Endocrinology 138:3940-3948(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-WISTAR; TISSUE-ISLETS OF LANGERHANS;
RA Svensson C.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U25680; AAB87095.1; -.
DR HSSP: P00740; 1IXA.
DR INTERPRO: IPR000152; -.
DR INTERPRO: IPR000561; -.
DR INTERPRO: IPR000742; -.
DR INTERPRO: IPR001881; -.
DR PFAM: PF00008; EGF_5.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_5.
DR PROSITE: PS01186; EGF_2; 6.
KW Glycoprotein.
SQ SEQUENCE 383 AA; 41253 MW; 7A6F90C02995C100 CRC64;

Query Match 12.3%; Score 82; DB 11; Length 383;
Best Local Similarity 31.1%; Pred. No. 0.11;
Matches 19; Conservative 9; Mismatches 21; Indels 12; Gaps 4;

QY 35 LKESCIRNDCETGCCORAPDNCEHSCAEGSGSLCQTQVFFGQYRAC---PCLRNLTG 91
DB 53 LCEKCVTSPGCVNGLCEE-PWQC---VCKEGWDGKFCFCEIDI-----RACTSTPCANNNGTC 103

QY 92 I 92
DB 104 V 104

RESULT 8
ID Q9YHU2 PRELIMINARY; PRT; 1254 AA.
AC Q9YHU2;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE SERRATEB.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi.
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RA Haddon C., Jiang Y.-J., Smithers L., Lewis J.;
RT "Delta-Notch signalling and the patterning of sensory cell

```

RT differentiation in the zebrafish ear: evidence from the mind bomb
 mutant.";
 RL Development 125:4637-4644(1998).
 DR EMBL; AF090432; AAC98354.1; -;
 DR HSSP; P00740; 1EDM.
 DR INTERPRO; IPR000152; -;
 DR INTERPRO; IPR000561; -;
 DR INTERPRO; IPR000742; -;
 DR INTERPRO; IPR000950; -;
 DR INTERPRO; IPR001438; -;
 DR INTERPRO; IPR001774; -;
 DR INTERPRO; IPR001881; -;
 DR PFAM; PF00008; EGF_15.
 DR PFAM; PF01414; DSL_1.
 DR PRINTS; PR00010; EGFLOOD.
 DR PROSITE; PS00010; ASX-HYDROXYL; UNKNOWN_7.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_16.
 DR PROSITE; PS00241; RECEPTOR_CYTOKINES_1; UNKNOWN_1.
 DR PROSITE; PS01186; EGF_2; 11.
 DR PROSITE; PS01187; EGF_CA; 6.
 DR Glycoprotein; EGF-like domain.
 KW Glycoprotein; EGF-like domain.
 SQ SEQUENCE 1254 AA; 137685 MW; BB902C35D3502B70 CRC64;

Query Match 12.2%; Score 81; DB 13; Length 1254;
 Best Local Similarity 28.4%; Pred. No. 0.41;
 Matches 23; Conservative 9; Mismatches 35; Indels 14; Gaps 4;
 QY 27 PTYNL-----LEKE-SCIRNQCETCCORAPDNCESHCAEKGSGSLCQTV 75
 Db 320 PDEVNCAPEGYSGKNCIEAHACVSNPCANGGTCEVTGTFCHC-PPGWEGPTCAKDM 378
 QY 76 FFGQYRACPLRLNLTCTYSKN 96
 Db 379 --DECASSPCAQGGTCIDLEN 397

RESULT 9
 ID Q9QXU5 PRELIMINARY; PRT; 107 AA.
 AC Q9QXU5;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE BV8 VARIANT 3 PRECURSOR.
 GN BV8.
 OS Mus. musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC STRAIN=129/SV; TISSUE=TESTIS;
 RA Wechselsberger C., Puglisi R., Lepperdinger G., Boitani C., Kreil G.;
 RT "The mammalian homologue of Bv8 from frog skin is mainly expressed in
 spermatoocytes.";
 RL FEBS Lett. 462:177-181(1999).
 DR EMBL; AF182066; AAF15261.1; -;
 DR HSSP; P25687; 11MT.
 KW Signal.
 FT SIGNAL
 SQ SEQUENCE 107 AA; 11556 MW; 8FB79BE4D4357C70 CRC64;

Query Match 12.1%; Score 80.5; DB 11; Length 107;
 Best Local Similarity 25.08%; Pred. No. 0.049;
 Matches 26; Conservative 16; Mismatches 47; Indels 15; Gaps 4;
 QY 8 LLLFLFFFLRGLSPTKYNLLEKESCIRNQCETGCCORAPDNCES--HCAEKG 65
 Db 9 LLLLLLLPLLF-----TPAGDAAVITGACDKDSQCGGMCACVSIWVKSIRICTPMG 61
 QY 66 SEGSLCQTV----FFQO--YRACPLRLNLTCTYSKNEKWLISIA 103

Db 62 QVGDSCHPLTRKVPFWGRMHHTCPLPLGLACLRTSFNRFLCA 105
 RESULT 10
 ID Q45448 PRELIMINARY; PRT; 126 AA.
 AC Q45448; O45598;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-JAN-1999 (Tremblrel. 09, Last annotation update)
 DE F35C5.12 PROTEIN.
 GN F35C5.12.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smye R.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Jier M., Johnston L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinon-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.

Query Match 12.1%; Score 80.5; DB 5; Length 126;
 Best Local Similarity 23.08%; Pred. No. 0.057;
 Matches 31; Conservative 13; Mismatches 44; Indels 47; Gaps 6;
 QY 9 LLLFLFFFLRGLSPTKYNLLEKESC--IRNQCETGC----- 49
 Db 1 MLFL--FFAFLL--SLQP-----FVRGACTFLDQACDEVCKTDFSEWYHGNGWDGFN 50
 QY 50 -----CQAPDNCESHCAEKGSEGLC--QTVFVGQYRACPLRLNLTCT 91
 Db 51 FSCCKVEYITPLDGSICETROMACSEKCKDQSGEGGCFPQLDSSHKSLRTACECFKQV 110
 QY 92 IYSKNEKWLISIAVGR 106
 Db 111 LRRKRSIQKRSYKR 125

Query Match 12.1%; Score 80.5; DB 5; Length 126;
 Best Local Similarity 23.08%; Pred. No. 0.057;
 Matches 31; Conservative 13; Mismatches 44; Indels 47; Gaps 6;
 QY 9 LLLFLFFFLRGLSPTKYNLLEKESC--IRNQCETGC----- 49
 Db 1 MLFL--FFAFLL--SLQP-----FVRGACTFLDQACDEVCKTDFSEWYHGNGWDGFN 50
 QY 50 -----CQAPDNCESHCAEKGSEGLC--QTVFVGQYRACPLRLNLTCT 91
 Db 51 FSCCKVEYITPLDGSICETROMACSEKCKDQSGEGGCFPQLDSSHKSLRTACECFKQV 110
 QY 92 IYSKNEKWLISIAVGR 106
 Db 111 LRRKRSIQKRSYKR 125

RESULT 11
 ID Q9U0E2 PRELIMINARY; PRT; 585 AA.
 AC Q9U0E2;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE SIGNAL RECEPTOR PROTEIN (FRAGMENT).

GN NOTCH.
 OS Tribolium castaneum (Red flour beetle).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
 OC Cucujiformia; Tenebrionidae; Tribolium.
 OX NCBI_TaxID=7070;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Tautz D., Lardelli M., Westin J., Tamme R.;
 RA "Embryonic expression of Tribolium Notch";
 RT Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AJ005083; CAB65469.1; -;
 DR HSP; P00740; IEDM.
 DR INTERPRO: IPR000083; -;
 DR INTERPRO: IPR000152; -;
 DR INTERPRO: IPR000561; -;
 DR INTERPRO: IPR000742; -;
 DR INTERPRO: IPR000800; -;
 DR INTERPRO: IPR001010; -;
 DR INTERPRO: IPR001438; -;
 DR INTERPRO: IPR001881; -;
 DR INTERPRO: IPR002049; -;
 DR PFAM; PF00008; EGF_11.
 DR PFAM; PF00066; notch; 2.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGFBLAMININ.
 DR PRINTS; PR00012; FNTYPEI.
 DR PRINTS; PR00287; THIONIN.
 DR PROSITE; PS00010; ASX_HYDROXYL; 6.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_12.
 DR PROSITE; PS01186; EGF_2; 9.
 DR PROSITE; PS01187; EGF_CA; 4.
 KW Receptor.
 FT NON_TER 1
 FT NON_TER 585
 FT NON_TER 585
 SQ SEQUENCE 585 AA; 63337 MW; 4CF7A51D0820D048 CRC64;

Query Match 12.08; Score 80; DB 5; Length 585;
 Best Local Similarity 25.8%; Pred. No. 0.27;
 Matches 17; Conservative 16; Mismatches 25; Indels 8; Gaps 3;
 QY 32 LLEKESICR-----NDCEGTCORAPDNCESHCSEKSGSLCQTQVFFGQYRACPC 86
 Db 138 MVSKDAALRKXVPLKLCNNGTCEDIGNSHRCHSD-GYSGSYCQFEI--NECDAPCQ 194
 QY 87 RNLFCI 92
 Db 195 NGGTCL 200

RESULT 12
 Q9XWD6 PRELIMINARY; PRT; 1111 AA.
 AC Q9XWD6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE Y47H9C.4 PROTEIN.
 GN Y47H9C.4
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=94150718; PubMed=7906398;
 RA Harris B.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Latreille N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Snelson N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkison-Sproat J., Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL; AL032657; CAA21739.1; -;
 DR HSP; P10968; 7MGA.
 DR INTERPRO: IPR000561; -;
 DR INTERPRO: IPR002049; -;
 DR PFAM; PF00008; EGF_11.
 DR PRINTS; PR00011; EGFBLAMININ.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_15.
 DR PROSITE; PS01186; EGF_2; 11.
 SQ SEQUENCE 1111 AA; 118803 MW; A39F374C008F9874 CRC64;

Query Match 12.08; Score 80; DB 5; Length 1111;
 Best Local Similarity 31.2%; Pred. No. 0.48;
 Matches 24; Conservative 8; Mismatches 27; Indels 18; Gaps 6;
 QY 44 DCETGC-CORAP--DNCESHC-AEKSGSLCQTQV---FFGQ--YRACPCILNLTCTIYS 94
 Db 244 ECKFECCNGCATCDNTNGKICKSGYHGALCENECVSFFGSGCTKCCLNQCDDSS 303
 QY 95 KNEKWLSTAYGRCKIG 111
 Db 304 S-----GECKCIG 311

RESULT 13
 Q28625 PRELIMINARY; PRT; 474 AA.
 AC Q28625;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE TUBULOINTERSTITIAL NEPHRITIS ANTIGEN.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=KIDNEY CORTEX;
 RC MEDLINE=95332335; PubMed=7608193;
 RA Nelson T.R., Charonis A.S., McIvor R.S., Butkowski R.J.;
 RT "Identification of a cDNA encoding tubulointerstitial nephritis
 antigen.";
 RL J. Biol. Chem. 270:16265-16270(1995).
 DR EMBL; U24270; AAC48477.1; -;
 DR HSP; P00787; 1THE.
 DR INTERPRO: IPR000169; -;
 DR INTERPRO: IPR000668; -;
 DR INTERPRO: IPR001212; -;
 DR PFAM; PF00112; Peptidase_C1; 2.
 DR PFAM; PF01033; Somatomedin_B; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 SQ SEQUENCE 474 AA; 54521 MW; 46AFA48FB7C29D1B CRC64;

Query Match 11.6%; Score 77.5; DB 6; Length 474;
 Best Local Similarity 25.0%; Pred. No. 0.43;
 Matches 28; Conservative 18; Mismatches 43; Indels 23; Gaps 7;

CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF 6-
CC 20 KDA, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 KDA).
CC EMBL: X63338; CAA44938.1; -.
CC INTERPRO: IPR002494; -.
CC PFAM: PF01500; Keratin_B2; 1.
KW Keratin.
SQ SEQUENCE 175 AA; 18218 MW; 0CB8BEFC9704837E CRC64;

Query Match 11.6%; Score 77; DB 4; Length 175;
Best Local Similarity 44.4%; Pred. No. 0.2;
Matches 16; Conservative 3; Mismatches 11; Indels 6; Gaps 2;
QY 38 SCIRNQDCETGCCQAPDNCESHCAEKSGSLCQT 73
DB 23 SCCQSPCCETSCCQ--PSCCETSCCQ-----SCCQT 52

Search completed: March 24, 2001, 13:25:43
Job time: 249 sec

QY 6 WLLLLFFLLFFELP-----LLTRGSLSPTKYNLE---LKESCIRNQDCET-GCC 50
DB 2 WTGYKKEFFIYLAKEIWTQKQKQVDLASYSRNHSILEGTRFKRSIFQGYCRSLGCC 61
QY 51 QRAPDNCESHCAEKSGSLCQTVFFGQYRA--CPCLRNLCIYKNEKWL 100
DB 62 EGRNDNCVTFQYE---ANALCYCKFCERENSDCCPDYKSF-C--QEEKEWL 107

RESULT 14
Q49549 PRELIMINARY; PRT: 1187 AA.
AC Q49549;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TremBLrel. 08, Last annotation update)
DE REPEAT REGIONS IN POTENTIAL METAL BINDING PROTEIN GENE REGION.
OS Mycoplasma hyorhinis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2100;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GD11;
RX MEDLINE=95014025; PubMed=7928953;
RA Deng G., McIntosh M.A.;
RT "An amplifiable DNA region from the Mycoplasma hyorhinis genome."
RL J. Bacteriol. 176:5929-5937(1994).
DR EMBL: L11447; AAA62228.1; -.
SQ SEQUENCE 1187 AA; 135027 MW; 66A82AE5B0EDE93E CRC64;

Query Match 11.6%; Score 77.5; DB 2; Length 1187;
Best Local Similarity 36.2%; Pred. No. 1;
Matches 25; Conservative 7; Mismatches 26; Indels 11; Gaps 6;
QY 27 PTKYNLLKESCIRNQDCETGCCQAPDNCESHCAEK--GSEGLCQ-TQVFFG-QYRA 82
DB 339 FGCGLKETEESC-----DCEACKQCEECNCS--CSELTGCGQCATSCAQEHCGCQEE 392
QY 83 CPCLRNLC 91
DB 393 CAC-PNTTC 400

RESULT 15
Q07628 PRELIMINARY; PRT: 175 AA.
AC Q07628;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)
DE KERATIN.
GN HB2B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93062886; PubMed=1279380;
RA Zhunabaeva B.D., Gening L.V., Gazaryan K.G.;
RT "[Cloning and structural characteristics of human hair keratin genes
RT rich in sulfur].";
RL Mol. Biol. (Mosk) 26:550-555(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93062886; PubMed=1279380;
RA Zhunabaeva B.D., Gening L.V., Gazaryan K.G.;
RT "[Cloning and structural characteristics of human hair keratin genes
RT rich in sulfur].";
RL Mol. Biol. (Mosk) 26:813-820(1992).

THIS PAGE BLANK (USPTO)

A:Accession: A42568
A:Molecule type: DNA
A:Residues: 1-112 <SIM>
A:Cross-references: GB:M95529; NID:g180842; PIDN:AA05818.1; PID:g1483624
A:Note: Sequence extracted from NCBI backbone (NCBIN:110576, NCBIN:110580)
R:Lowe, M.E.; Rosenblum, J.L.; McEwen, P.; Strauss, A.W.
Biochemistry 29, 823-828, 1990
A:Title: Cloning and characterization of the human colipase cDNA.
A:Reference number: A33949; MUID:90248429
A:Accession: A33949
A:Molecule type: mRNA
A:Residues: 1-112 <LOW>
A:Cross-references: GB:J02883; NID:g180885; PIDN:AAA52054.1; PID:g180886
A:Note: Evidence of partial N-glycosylation, possibly at Asn-43
R:Sternby, B.; Engstrom, A.; Hellman, U.; Vihert, A.M.; Sternby, N.H.; Borgstrom, B.
Biochim. Biophys. Acta 784, 75-80, 1984
A:Title: The primary sequence of human pancreatic colipase.
A:Reference number: A90652; MUID:84104937
A:Accession: A03163
A:Molecule type: protein
A:Residues: 23-108 <STE>
C:Comment: Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms a 1:1 stoichiometric complex with the enzyme, which are known to have an inhibitory effect on the enzyme is washed off by bile salts, which are known to have an inhibitory effect on the enzyme.
C:Genetics:
A:Gene: GDB:CLPS
A:Cross-references: GDB:127277; OMIM:120105
A:Map position: 6pter-6p21.1
A:Introns: 28/3; 69/3
C:Superfamily: colipase
C:Keywords: lipid digestion; lipid hydrolysis; pancreas
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-22/Domain: amino-terminal propeptide #status predicted <APP>
F:23-108/Product: colipase #status experimental <MAT>
F:109-112/Domain: carboxyl-terminal propeptide #status predicted <CPP>
F:34-104,40-56,44-80,45-78,66-86/Disulfide bonds: #status predicted
F:69,72,75,76/Binding site: micellar substrate (Lys, Tyr, Tyr, Tyr) #status predicted

Query Match 20.3%; Score 135.5; DB 1; Length 112;
Best Local Similarity 28.8%; Pred. No. 6.8e-06;
Matches 32; Conservative 17; Mismatches 51; Indels 11; Gaps 3;

QY 12 LFFFLLELLRGLSPKYNL--LELKESCIRNQDCTGCCORAPDNCESHCARKEGSEG 68
Db 4 ILILLVSLVAVAAPGRGIIINLENGELOMNSAQCKNSCCOHSSALGLARTSTWSEN 63

QY 69 SLCTQVFFGQYRACPLRLNLTCIYSKNEKWL-----SIAYGRCKQIGRK 114
Db 64 SECSVKTLGIYKPCERGLTC---EGDKTIVGSIITNTNFGICHIDAGRSK 111

RESULT 3
S53510
pancreatic colipase - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 17-Mar-1999
C:Accession: S53510
R:Rugani, N.; Carrier, F.; Thim, L.; Borgstrom, B.; Sarda, L.
Biochim. Biophys. Acta 1247, 185-194, 1995
A:Title: Lipid binding and activating properties of porcine pancreatic colipase split at the activation site.
A:Reference number: S53510; MUID:95210311
A:Accession: S53510
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-95 <RUG>
C:Superfamily: colipase

Query Match 20.1%; Score 134; DB 2; Length 95;
Best Local Similarity 34.1%; Pred. No. 8.3e-06;
Matches 29; Conservative 11; Mismatches 37; Indels 8; Gaps 2;

QY 33 LELEKESCIRNQDCTGCCORAPDNCESHCARKEGSEGSLCQTQVFFGQYRACPLRLNLTCI 92

Db 11 LDEGELCLNSAQCKNSCCQHDHTILSRCAKARENSECSAFTLYGVYKPCERGLTC- 69
QY 93 YSKNEKWL-----SIAYGRCKQIGR 112
Db 70 --EGDKSLVGSITNTNFGICHINVR 92

RESULT 4
XLHOB
colipase B precursor - horse
N:Alternate names: procolipase B
C:Species: Equus caballus (domestic horse)
C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 26-Apr-1996
C:Accession: A03165; B90220
R:Bonicel, J.; Couchoud, P.; Foglizzo, E.; Desnuelle, P.; Chapus, C.
Biochim. Biophys. Acta 669, 39-45, 1981
A:Title: Amino acid sequence of horse colipase B.
A:Reference number: A90637; MUID:82046794
A:Accession: A03165
A:Molecule type: protein
A:Residues: 1-96 <BON>
R:Julien, R.; Bechis, G.; Gregoire, J.; Rathelot, J.; Rochat, H.; Sarda, L.
Biochem. Biophys. Res. Commun. 95, 1245-1252, 1980
A:Title: Evidence for the existence of two isocolipases in horse pancreas.
A:Reference number: A90220; MUID:81021166
A:Accession: B90220
A:Molecule type: protein
A:Residues: 1-21, 'E', '23-28, 'T', '30-55 <JUL>
C:Comment: Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms a 1:1 stoichiometric complex with the enzyme, which are known to have an inhibitory effect on the enzyme is washed off by bile salts, which are known to have an inhibitory effect on the enzyme.
C:Superfamily: colipase
C:Keywords: lipid digestion; lipid hydrolysis; pancreas
F:1-3/Domain: propeptide #status experimental <PRO>
F:16-96/Product: colipase B #status experimental <MAT>
F:17-87,23-39,27-63,28-61,49-69/Disulfide bonds: #status predicted
F:52,55,58,59/Binding site: micellar substrate (Trp, Tyr, Tyr, Tyr) #status predicted

Query Match 20.1%; Score 134; DB 1; Length 96;
Best Local Similarity 39.7%; Pred. No. 8.3e-06;
Matches 25; Conservative 6; Mismatches 32; Indels 0; Gaps 0;

QY 33 LELEKESCIRNQDCTGCCORAPDNCESHCARKEGSEGSLCQTQVFFGQYRACPLRLNLTCI 92
Db 11 LEAGEITMNSAQCKNSCCORAPDNCESHCARKEGSEGSLCQTQVFFGQYRACPLRLNLTCI 70

QY 93 YSK 95
Db 71 VDK 73

RESULT 5
XLHOA
colipase A precursor - horse
N:Alternate names: procolipase A
C:Species: Equus caballus (domestic horse)
C:Date: 14-Nov-1983 #sequence_revision 04-Dec-1986 #text_change 26-Apr-1996
C:Accession: A03164; A91119; A90220
R:Sternby, B.; Engstrom, A.; Hellman, U.; Vihert, A.M.; Sternby, N.H.; Borgstrom, B.
Biochim. Biophys. Acta 784, 75-80, 1984
A:Title: The primary sequence of human pancreatic colipase.
A:Reference number: A90652; MUID:84104937
A:Accession: A03164
A:Molecule type: protein
A:Residues: 1-96 <STE>
A:Note: residues 56-59 were positioned by homology; no overlap was obtained for 65-66
R:Pierrot, M.; Astier, J.P.; Astier, M.; Charles, M.; Drenth, J.
Eur. J. Biochem. 123, 347-354, 1982
A:Title: Pancreatic colipase: crystallographic and biochemical aspects.
A:Reference number: A91119; MUID:82186702
A:Accession: A91119
A:Molecule type: protein

A:Accession: A03162
A:Molecule type: protein
A:Residues: 1-95 <STE>
R:Charles, M.; Erlanson, C.; Blanchetta, J.: Joffre, J.; Guidoni, A.: Roversy, M.
Biochim. Biophys. Acta 359, 186-197, 1974

A:title: The primary structure of porcine colipase II. I. The amino acid sequence.

A:Reference number: A90593; MUID:74290109

A:Accession: A90593

A:Molecule type: protein

A:Residues: 6-91 <CHA>

R:Erlanson, C.: Charles, M.; Astier, M.; Desnuelle, P.
Biochim. Biophys. Acta 359, 198-203, 1974

A:title: The primary structure of porcine colipase II. II. The disulfide bridges.

A:Reference number: A90594; MUID:74290110

A:Contents: annotation; disulfide bonds

C:Comment: Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms a 1:1 s
se the enzyme is washed off by bile salts, which are known to have an inhibitory effe

C:Comment: Residues 6-9 and Arg-92 are considered essential for the function of collip

C:Superfamily: colipase

C:Keywords: lipid digestion; lipid hydrolysis; pancreas

F:1-5/Domain: propeptide #status experimental <PRO>

F:6-9/Product: colipase II #status experimental <MAT>

E:17-87, 23-39, 27-63, 28-61, 49-69/Disulfide bonds: #status experimental

F:52, 55, 58, 59/Binding site: micellar substrate (Phe, Tyr, Tyr) #status predicted

Query Match 19.2% Score 128; DB 1; Length 95;
Best Local Similarity 32.9%; Pred. No. 3e-05;
Matches Conservative 11; Mismatches 38; Indels 8; Gaps 2;

QY 33 LELKESCI RNDCETGCCORAPDNCESHCAFKGSGSLCQTQVFVGQRACPLRNITCI 92
 | : | : | : ||| || | | | | : | | | |
Db 11 LDDELCLNSAQSKNCCQHDTILLLRCALKARENSECSAFITYGVVVKPCBRLTCT - 69

QY 93 YSNKEKWL-----SIAYGRCKQIGR 112
 | : | : | : ||| :

Dd 70 --EGDKSIVGSITNTNFGLHNVR 92
 | : | : | : ||| :

RESULT 8
I51909
Colipase precursor - rat
N:Alternate names: procollipase
C:Species: Rattus norvegicus (Norway rat)
C>Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 20-Apr-2000
C:Accession: I51909; A34623
R:Payne, R.M.; Sims, H.F.; Jennens, M.L.; Lowe, M.E.
Am. J. Physiol. 266, G914-G921, 1994
A:title: Rat pancreatic lipase and two related proteins: enzymatic properties and mRNA
A:Reference number: I51909; MUID:94262798
A:Accession: I51909
A>Status: preliminary; translated from GB/EMBL/DDDBJ
A:Molecule type: mRNA
A:Residues: 1-112 <PAY>
A:CROSS-references: GB:M58370; NID:g203504; PIDN:AAA20505.1; PID:g203505
R:Wicker, C.: Pulgoverner, A.
Biochem. Biophys. Res. Commun. 167, 130-136, 1990
A:title: Rat pancreatic colipase mRNA: nucleotide sequence of a cDNA clone and nutritio
A:Reference number: A34623; MUID:90179738
A:Accession: A34623
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-17,'v','l9-112 <WTc>
A:CROSS-references: GB:M33333; NID:g203502; PIDN:AAA40943.1; PID:g203503
C:Superfamily: colipase
C:Keywords: lipid digestion; lipid hydrolysis; pancreas
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-112/Product: colipase #status predicted <MAT>

Query Match 18.3% Score 122; DB 2; Length 112;
Best Local Similarity 31.9%; Pred. No. 0.00012;
Matches Conservative 8; Mismatches 38; Indels 16; Gaps 2;

```
QY 33 LEKESCIRNQDCTGCCORAPDNCESHAEGSGSLCQTVFFGQYRACPLRLNLTG- 91
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 28 LEDGEICVNSMOCKSRCCQHDHILGARTKHAMENSECSPKTLGYIYRCPCERGLTCE 87
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 92 -----IYSKNEKWLSTAYGRCKQKIG 114
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 88 GDRSIIGAITNTN-----YGVCLDSTRSK 111
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

RESULT 9
S08220
platelet-derived growth factor chain A precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 16-Jul-1999
C:Accession: S08220
R:Bejcek, B.E.; Li, D.Y.; Deuel, T.F.
Nucleic Acids Res. 18, 680, 1990
A:Title: Nucleotide sequence of a cDNA clone of Xenopus platelet-derived growth factor A
A:Reference number: S08220; MUID: 90175018
A:Accession: S08220
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-215 <BEJ>
A:Cross-references: EMBL:X17545; NID:g64973; PIDN:CAA35583.1; PID:g64974
C:Superfamily: platelet-derived growth factor
C:Keywords: alternative splicing; growth factor
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-91/Domain: propeptide #status predicted <PRO>
F;92-215/Product: platelet-derived growth factor chain A #status predicted <MAT>

Query Match 12.5%; Score 83.5; DB 2; Length 215;
Best Local Similarity 22.8%; Pred. No. 0.74; Mismatches 48; Indels 19; Gaps 5;
Matches 26; Conservative 21;

QY 17 LFLLTRGSLSPTKYNLELKESIRNQDCTGCCORAPDNC-----SHCAEK-----G 65
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 107 IYEIPRSDIPTSANFL-IWPCVEVKRC-TGCCNTSSVKQPSRIHRSVKVAKVEYVR 164
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

QY 66 SEGSLCQTVFFGQYRACPLRLNLTCTIYSKNEKWLSTAYGRCKQKIGRKAKKM 119
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 165 KKPRLKVLVRLBEHLECTCTANSNDYREET-----GRTRETKRKRKKL 212
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

RESULT 10
T21762
hypothetical protein F35C5.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21762; T22937
R:Sims, M.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19469
A:Accession: T21762
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-126 <WTL>
A:Cross-references: EMBL:Z81076; NID:e1061694; PIDN:CAB03062.1; GSPDB:GN000020; CESP:F35C5
A:Experimental source: clone F35C5
R:Smey, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19639
A:Accession: T22937
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-126 <WTL>
A:Cross-references: EMBL:Z81556; PIDN:CAB04528.1; GSPDB:GN000020; CESP:F35C5.12
A:Experimental source: clone F58G1
C:Genetics:
A:Gene: CESP:F35C5.12
A:Map position: 2
A:Introns: 18/2; 52/2; 81/3
```

```
Query Match 12.1%; Score 80.5; DB 2; Length 136;
Best Local Similarity 23.0%; Pred. No. 0.95;
Matches 31; Conservative 13; Mismatches 44; Indels 47; Gaps 6;

QY 9 LFLFLFFFLTRGSLSPTKYNLELKESIRNQDCTGCC-----IRNODCTGC----- 49
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 1 MLFL--FFAFL--SLOP-----FVRGACTFLDHOACDEVCKTDSFYGCHCNGWDGFN 50
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

QY 50 -----CORAPDNCESHAEGSGSLC--QTVFFGQYRACPLRLNLTG 91
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 51 FSCKYEYITPLDGSICETROMACSEKDKGSGEGGFCFPQLDSHKSLRTACECFKKLQV 110
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

QY 92 IYSKNEKWLSTAYGR 106
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 111 LRRKRSIQKRSYKR 125
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

RESULT 11
T26972
hypothetical protein Y47H9C.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T26972
R:Harris, B.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20293
A:Accession: T26972
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1111 <WIL>
A:Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN000019; CESP:Y47H9C.4
A:Experimental source: clone Y47H9C
C:Genetics:
A:Gene: CESP:Y47H9C.4
A:Map position: 1
A:Introns: 50/2; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

Query Match 12.0%; Score 80; DB 2; Length 1111;
Best Local Similarity 31.2%; Pred. No. 5.3;
Matches 24; Conservative 8; Mismatches 27; Indels 18; Gaps 6;

QY 44 DCETGC-CQRAP--DNCESHC-AEKGSGSLCQTVQV---FFGQ--YRACPLRLNLTCTIYS 94
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 244 ECKFECCQNGATCDNTNGCKICKSGYHGALCENECVSGFFGSGCTQKCDCLNNQNCDS 303
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

QY 95 KNEKWLSTAYGRCKQIG 111
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 304 S-----GECKCIG 311
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

RESULT 12
I51550
platelet-derived growth factor A chain long form precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 16-Jul-1999
C:Accession: I51550
R:Mercola, M.; Melton, D.A.; Stiles, C.D.
Science 241, 1223-1225, 1988
A:Title: Platelet-derived growth factor A chain is maternally encoded in Xenopus embr
A:Reference number: I51550; MUID: 88321676
A:Accession: I51550
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-226 <MER>
A:Cross-references: GB:M23237; NID:g214648; PIDN:AAA49927.1; PID:g214649
C:Superfamily: platelet-derived growth factor

Query Match 11.9%; Score 79; DB 2; Length 226;
```

Best Local Similarity 21.8%; Pred. No. 2;
Matches 26; Conservative 22; Mismatches 53; Indels 18; Gaps 5;

QY 17 LFLTRGSLPTKYNLLEKESCIARNODCEGCGCORAPDNCE-----SHCAFK-----G 65
Db 107 IYEIPRQIDPISANFL-IWPCVEVARK-TGCCNTSVKCPRIHRSVKVAKVEYVR 164
QY 66 SEGSLCQTQVFFGQYRACPCPLRLNLTCTIYSKNEKL-----SIAYGRCKIGRQKLAKKM 119
Db 165 KKPALKVLRLHELECTCTANSNDYREBETGFTSPALVLTGRTRTGKKOKRKL 223

RESULT 13
S18188
notch protein homolog - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
C:Accession: S18188
R;Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 113, 199-205, 1991
A:Title: A homolog of Drosophila Notch expressed during mammalian development.
A:Reference number: S18188; MUID:92111383
A:Accession: S18188
A:Molecule type: mRNA
A:Residues: 1-2531 <WEI>
A:Cross-references: EMBL:X57405; NID:g57634; PID:g57635
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
F:987-1018/Domain: EGF homology <EGF1>
F:1025-1056/Domain: EGF homology <EGF2>
F:1233-1264/Domain: EGF homology <EGF3>
F:1917-1949/Domain: ankyrin repeat homology <AN1>
F:1950-1982/Domain: ankyrin repeat homology <AN2>
F:1984-2016/Domain: ankyrin repeat homology <AN3>
F:2017-2049/Domain: ankyrin repeat homology <AN4>
F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 11.8%; Score 78.5; DB 2; Length 2531;
Best Local Similarity 29.5%; Pred. No. 13;
Matches 23; Conservative 11; Mismatches 35; Indels 9; Gaps 5;

QY 15 FFLFLTRGSLPT-KYNLLEKESCIARNODCEGCGCORAPDNCESHCAKSGSLCQT 73
Db 624 YYLCLLKGTTGPNCETNL-----DDCASN-PCDSGTCLDKIDGYECAC-EPGYTGSMCNV 677

QY 74 QVFFGQYRACPCPLRLNLT 91
Db 678 NI--DECAGSPCHNGGTC 693

RESULT 14
S43910
gibberellin-regulated protein RSI-1 precursor - tomato
N:Alternate names: GAST1 protein homolog
C:Species: Lycopersicon esculentum (tomato)
C:Date: 13-Jan-1995 #sequence_revision 25-Apr-1997 #text_change 24-Sep-1999
C:Accession: S43910; S43909
R;Taylor, B.H.; Scheuring, C.F.
Mol. Gen. Genet. 243, 148-157, 1994
A:Title: A molecular marker for lateral root initiation: the RSI-1 gene of tomato (Lycopersicon esculentum)
A:Reference number: S43909; MUID:94232181
A:Accession: S43910
A:Molecule type: DNA
A:Residues: 1-96 <TAY>
A:Cross-references: EMBL:L22189; NID:g405586; PIDN:AAA20130.1; PID:g405587
A:Accession: S43909
A:Molecule type: mRNA
A:Residues: 1-96 <TAY>
A:Cross-references: EMBL:L22188; NID:g405584; PIDN:AAA20129.1; PID:g405585
C:Genetics:
A:Gene: RSI-1
A:Introns: 28/3; 37/1
C:Superfamily: gibberellin-regulated protein GASA2

F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-96/Product: gibberellin-regulated protein RSI-1 #status predicted <MAT>

Query Match 11.7%; Score 78; DB 2; Length 96;
Best Local Similarity 28.4%; Pred. No. 1.3;
Matches 25; Conservative 9; Mismatches 32; Indels 22; Gaps 4;

QY 10 LFLPFFFLTRGSLPTKYNLLEKESCIARNODCEGCGCORAPDN-----CESH 60
Db 11 LLLSMFLILLTFNSV-VEGYNKL-----RPTDCKPCTYRCSATSHKKKPCMEFFCQKC 62

QY 61 CAEKSGESGLCQTQVFFGQYRACPCPLRN 88
Db 63 CAT-----CLCPKGVYGNKQSCPCYNN 85

RESULT 15
A57480
tubulointerstitial nephritis antigen precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change 05-Nov-1999
C:Accession: A57480
R;Nelson, T.R.; Charonis, A.S.; McIvor, R.S.; Butkowski, R.J.
J. Biol. Chem. 270, 16265-16270, 1995
A:Title: Identification of a cDNA encoding tubulointerstitial nephritis antigen.
A:Reference number: A57480; MUID:95332335
A:Accession: A57480
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-474 <NEL>
A:Cross-references: GB:U24270; NID:g818410; PIDN:AAC48477.1; PID:g818411
C:Superfamily: somatomedin B homology
C:Keywords: basement membrane; extracellular matrix; glycoprotein
F:51-104/Domain: somatomedin B homology <SBH2>

Query Match 11.6%; Score 77.5; DB 2; Length 474;
Best Local Similarity 25.0%; Pred. No. 4.8;
Matches 28; Conservative 18; Mismatches 43; Indels 23; Gaps 7;

QY 6 WLLLLFLFFFLF-----LLTRGSLPTKYNLLE---LKESCIRNQDCET-GCC 50
Db 2 WTGYKFFIFFYLAKEIWTPEKQKQVDLASYSRRNHSILEGTRFKRSIFQGYCRSLGCC 61

QY 51 QRAPDNCESHCAKSGESGLCQTQVFFGQYRA--CPCLRLNLTCTIYSKNEKL 100
Db 62 EGRDNCVTQFYE---ANALCYCDKFCERENSDCCPDYKSF-C--QEEKEWL 107

Search completed: March 24, 2001, 13:23:16
Job time: 106 sec

QY 79 QYRACPLRLNLTCTYKNEKW--LSIAYGRCKIGROKLA 117
 Db 598 EKCMPECHTIPCLLSPWSEWSDCSVTGCKGMRT-RQRLMK 637

RESULT 2

US-08-313-288B-12
 ; Sequence 12, Application US/08313288B
 ; Patent No. 5750502
 ; GENERAL INFORMATION:
 ; APPLICANT: Jessell, Thomas M. and Avihu Klar
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
 ; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/313,288B
 ; FILING DATE: January 5, 1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0526
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 802 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US-08-313-288B-12

Query Match 11.4%; Score 76; DB 1; Length 802;
 Best Local Similarity 21.8%; Pred. No. 5.7;
 Matches 22; Conservative 17; Mismatches 44; Indels 18; Gaps 4;

QY 34 ELKESCI RNQDCETGCCORAP-----DNCESHCAE-----KGSEGLCQTQVFFG 78
 Db 538 ETEKCI VNEECSPSCLVTGWEWDECSASCTGMRHRMIKMT PADGSMCKAETTEA 597

RESULT 3

PCT-US93-03164-12
 ; Sequence 12, Application PC/TUS9303164
 ; GENERAL INFORMATION:
 ; APPLICANT: Jessell, Thomas M
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
 ; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham
 ; STREET: 30 Rockefeller Plaza
 ; CITY: New York

STATE: New York
 COUNTRY: USA
 ZIP: 10112
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/03164
 FILING DATE: 19930402
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 40028
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 977-9550
 TELEFAX: (212) 664-0525
 TELEX: 422523 COOP UI
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 802 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US93-03164-12

Query Match 11.4%; Score 76; DB 4; Length 802;
 Best Local Similarity 21.8%; Pred. No. 5.7;
 Matches 22; Conservative 17; Mismatches 44; Indels 18; Gaps 4;

QY 34 ELKESCI RNQDCETGCCORAP-----DNCESHCAE-----KGSEGLCQTQVFFG 78
 Db 538 ETEKCI VNEECSPSCLVTGWEWDECSASCTGMRHRMIKMT PADGSMCKAETTEA 597

RESULT 4

US-08-597-545-2
 ; Sequence 2, Application US/08597545
 ; Patent No. 5580738
 ; GENERAL INFORMATION:
 ; APPLICANT: LABORDA, Jorge
 ; TITLE OF INVENTION: Delta-Like Gene Expressed In
 ; TITLE OF INVENTION: Neuroendocrine Tumors
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington, D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/597,545
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/989,537
 ; FILING DATE: 11-DEC-1992
 ; ATTORNEY/AGENT INFORMATION:


```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

	Query Match	11.1%	Score 74;	DB 1;	Length 385;
	Best Local Similarity	31.1%;	Pred. No. 4.2;	Indels	Gaps
	Matches	19;	Conservative	7;	Mismatches 23;
QY	35	LKESCIRNQDCETGCCQAPDNCSEHCAEKGSEGLCQTQVFFQYRAC	----	PCLRNUTC	91
Db	53	LCDKCVTAPGCVNGVC-KEPWCQ	----	ICKDGDGKGFEDV	----
				RACTSTPCANNCTC	103

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

```
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-03164-10

Query Match 11.0%; Score 73; DB 4; Length 807;
Best Local Similarity 20.4%; Pred. No. 11;
Matches 21; Conservative 19; Mismatches 45; Indels 18; Gaps 4;

Qy 34 ELKESIRNQDCTGCCQRP-----DNCESHCA-----EKGSEGLCQTQVFFG 78
Db 543 EETKCTVNEECSPSLTEWGEWDGCSATCGMGKRRHVMKMSPADGSMCKRAETSOA 602

Qy 79 QYRACPLRLNLTCIYSKNEKW--LSIAYGRCKIGRQKLAKKM 119
Db 603 EKCMMPECHTIFCLLPWSEWSDCVTCGKGMRT-RQRLKSL 644

RESULT 13
US-08-036-555B-40
; Sequence 40, Application US/08036555B
; Patent No. 5530109
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,555B
; FILING DATE: 24-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.

; NAME: White, John P
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 5250.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-036-555B-40

Query Match 10.9%; Score 72.5; DB 1; Length 39;
Best Local Similarity 31.9%; Pred. No. 0.57;
Matches 15; Conservative 2; Mismatches 19; Indels 11; Gaps 2;

Qy 45 CETGCCQRPDNCESHAKEGSEGLCQTQVFFGOYRACPLRLNLT 91
Db 1 CTCGCCCKCRRTTCACRCA--GAAGGTCTT-----CTCCTTCTC 36

RESULT 14
US-08-469-569-40
; Sequence 40, Application US/08469569
; Patent No. 5606032
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,569
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 5250.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
```


us-09-215-435-179.ra1

Sat Mar 24 13:26:25 2001

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	154	28.6	110	1	IBP_CARCR	P00993	caretta car
2	134	24.9	352	1	AMP_BOVIN	P00978	bos taurus
3	130	24.1	62	1	IP52_ANESU	P10280	anemonia su
4	130	24.1	83	1	ELAC_WACEU	O62845	macropus eu
5	129.5	24.0	352	1	AMP_HUMAN	P02760	homo sapien
6	127	23.6	65	1	IVB3_VIPAA	P00992	vipera ammo
7	127	23.6	123	1	IATR_HORSE	P04365	equus cabal
8	127	23.6	123	1	IATR_SHEEP	P13371	ovis aries
9	126	23.4	67	1	IBPC_BOVIN	P00976	bos taurus
10	126	23.4	102	1	ELAC_TRIUV	Q28143	trichosurus
11	124	23.0	235	1	TFP2_HUMAN	P48307	homo sapien
12	124	23.0	349	1	AMP_HOUSE	Q07456	mus musculu
13	123	22.8	60	1	IVB1_DENPO	P00979	dendroaspis
14	122	22.6	349	1	AMP_RAT	Q64240	rattus norv
15	121	22.4	300	1	TFPI_RABIT	P19761	oryctolagus
16	120	22.3	304	1	TFPI_MACMU	Q28864	macaca mula
17	118	21.9	62	1	IVBT_ERIMA	P24541	eristocophi
18	118	21.9	337	1	AMP_PIG	Q34366	sus scrofa
19	118	21.9	346	1	AMP_MERUN	Q62577	meriones un
20	117	21.7	507	1	SBPI_SARBU	P26228	sarcophaga
21	117	21.7	302	1	TFPI_RAT	Q02445	rattus norv
22	116	21.5	304	1	TFPI_HUMAN	P10646	homo sapien
23	115	21.3	59	1	IVB1_DENAN	P00980	dendroaspis
24	113	21.0	349	1	AMP_MESAU	Q60559	mesocricetu
25	112.5	20.9	355	1	AMP_PLEPL	P36992	pleuronecte
26	111	20.6	57	1	IVBC_NAJNA	P19859	naja naja
27	110	20.4	2944	1	CA17_HUMAN	Q02388	homo sapien
28	109	20.2	60	1	TXCA_DENAN	P81658	dendroaspis
29	109	20.2	164	1	TKD1_BOVIN	Q82801	bos taurus
30	108.5	20.1	765	1	APR2_RAT	P15943	rattus norv
31	108	20.0	62	1	ISCC_BOMMO	P10832	bombux mori
32	108	20.0	65	1	IVB1_BUNFA	P25660	hungarus fa
33	107.5	19.9	763	1	APR2_HUMAN	Q06481	homo sapien

RESULT 2
AMBP_BOVIN

ID AMBP_BOVIN STANDARD; PRT; 352 AA.

AC P00978; P35420; Q28020;

AD 21-JUL-1986 (Rel. 01, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE AMBP PROTEIN PRECURSOR (CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE TRYPsin INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30) (BI-14)
DE (CUMULUS EXTRACELLULAR MATRIX STABILIZING FACTOR) (ESF)]).

GN AMBP OR ITIL.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.

[1]

RN SEQUENCE FROM N.A.

RP TISSUE=LIVER;

RC MEDLINE=96201710; PubMed=8611630;

RX Lindqvist A., Aakerstroem B.;

RA "Bovine alpha 1-microglobulin/Bikunin. Isolation and characterization
RT of liver cDNA and urinary alpha 1-microglobulin.";

RT Biochim. Biophys. Acta 1306:98-106(1996).

RL [2]

RN SEQUENCE OF 227-349.

RP MEDLINE=85225967; PubMed=2408637;

RC Hochstrasser K., Wachter E., Albrecht G.J., Reisinger P.;

RX "Kunitz-type proteinase inhibitors derived by limited proteolysis of
RA the inter-alpha-trypsin inhibitor, X. The amino-acid sequences of the
RT trypsin-released inhibitors from horse and pig inter-alpha-trypsin
RT inhibitors.";

RL Biol. Chem. Hoppe-Seyler 366:473-478(1985).

[3]

RN SEQUENCE OF 227-348.

RP MEDLINE=84133807; PubMed=6199275;

RC Hochstrasser K., Wachter E.;

RX "Kunitz-type proteinase inhibitors derived by limited proteolysis of
RA the inter-alpha-trypsin inhibitor, VII. Determination of the
RT amino-acid sequence of the trypsin-released inhibitor from bovine
RT inter-alpha-trypsin inhibitor.";

RL Hoppe-Seyler's Z. Physiol. Chem. 364:1679-1687(1983).

[4]

RN SEQUENCE OF 206-219.

RP TISSUE=FETAL SERUM;

RC MEDLINE=92291130; PubMed=1376324;

RX Chen L., Mao S.-J.T., Larsen W.J.;

RA "Identification of a factor in fetal bovine serum that stabilizes the
RT cumulus extracellular matrix. A role for a member of the inter-alpha-
RT trypsin inhibitor family.";

RL J. Biol. Chem. 267:12380-12386(1992).

[5]

RN REACTIVE SITES.

RP MEDLINE=84133808; PubMed=6199276;

RC Hochstrasser K., Albrecht G.J., Schoenberger O.L., Wachter E.;

RX "Kunitz-type proteinase inhibitors derived by limited proteolysis of
RA the inter-alpha-trypsin inhibitor, VII. Characterization of the
RT bovine inhibitor as double-headed trypsin-elastase inhibitor.";

RL Hoppe-Seyler's Z. Physiol. Chem. 364:1689-1696(1983).

CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC AND ALBUMIN.

-1- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYOSOMAL GRANULOCYTIC
CC ELASTASE.

-1- FUNCTION: MAY DIFFUSE INTO FOLLICULAR FLUID AFTER AN OVULATORY
CC STIMULUS TO ACT AS STRUCTURAL LINKER THAT ENSURE NORMAL CUMULUS
CC EXPANSION, THROUGH STABILIZATION OF THE CUMULUS EXTRACELLULAR
CC MATRIX THUS SUPPORTING THE PROCESS OF OVULATION.

-1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC SEPARATELY FUNCTIONING PROTEINS.

```
CC CC -!- SUBCELLULAR LOCATION: SECRETED.
```

```
CC CC -!- TISSUE SPECIFICITY: FOUND IN THE WHEY FRACTION OF MILK.
```

```
CC CC -!- DEVELOPMENTAL STAGE: EXPRESSED ONLY IN THE EARLY STAGE OF
```

```
CC CC LACTATION. DECREASE IN EXPRESSION CORRELATES WITH A CHANGE IN THE
```

```
CC CC SUCKLING PATTERN OF THE YOUNG.
```

```
CC CC -!- PTM: N-GLYCOSYLATED (PROBABLE).
```

```
CC CC -!- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
```

```
-----
```

```
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
```

```
CC CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC CC use by non-profit institutions as long as its content is in no way
```

```
CC CC modified and this statement is not removed. Usage by and for commercial
```

```
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC CC or send an email to licenselib@sib.ch).
```

```
-----
```

```
CC CC EMBL; AJ000490; CAA04128.1; -;
```

```
DR DR INTERPRO; IPR002223; -;
```

```
DR DR PFAM; PF00014; Kunitz-BPTI; 1.
```

```
DR DR PRINTS; PR00759; BASICPTASE.
```

```
DR DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
```

```
KW Serine protease inhibitor; Glycoprotein; Lactation.
```

```
FT DISULFID 23 73 BY SIMILARITY.
```

```
FT FT DISULFID 32 56 BY SIMILARITY.
```

```
FT FT DISULFID 48 69 BY SIMILARITY.
```

```
FT ACT_SITE 33 34 REACTIVE_BOND_BY_SIMILARITY.
```

```
FT DOMAIN 76 83 POLY-ASN.
```

```
FT CARBOHYD 14 14 N-LINKED_GLCNAC... (POTENTIAL).
```

```
FT FT CARBOHYD 31 31 N-LINKED_GLCNAC... (POTENTIAL).
```

```
FT FT CARBOHYD 42 42 N-LINKED_GLCNAC... (POTENTIAL).
```

```
SQ SEQUENCE 83 AA; 9583 MW; CD94CD35EF3175E1 CRC64;
```

```
Query Match 24.1%; Score 130; DB 1; Length 83;
```

```
Best Local Similarity 56.1%; Pred. No. 3.7e-08;
```

```
Matches 23; Conservative 4; Mismatches 14; Indels 0; Gaps 0;
```

```
QY 41 CKLDNFGSGYEYHFRRFYFNRTSKRCETFFVFSGCNGNLNNF 81  
| : | : | : | : | : | : | : | : | : |  
DB 23 CLLPVRGNCGSSQLHLHYNTTSCTCFYISGCNGNRNNF 63
```

```
RESULT 5
```

```
ID AMBP_HUMAN STANDARD; PRT; 352 AA.
```

```
AC P02760; P02759; P00977;
```

```
DT 21-JUL-1986 (Rel. 01, Created).
```

```
DT 13-AUG-1987 (Rel. 05, Last sequence update)
```

```
DE 30-MAY-2000 (Rel. 39, Last annotation update)
```

```
DE AMBP PROTEIN PRECURSOR [CONTAINS: ALPHA-1-MICROGLOBULIN (PROTEIN HC)
```

```
DE (COMPLEX-FORMING GLYCOPROTEIN HETEROGENEOUS IN CHARGE); INTER-ALPHA-
```

```
DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-IC) (BIKUNIN) (HI-30)].
```

```
GN AMBP OR ITIL OR HCP.
```

```
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
```

```
RN [1]
```

```
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=91214554; PubMed=1708673;
```

```
RA Vetr H., Gebhard W.;
```

```
RL "Structure of the human alpha 1-microglobulin-bikunin gene.";
```

```
RN Biol. Chem. Hoppe-Seyler 371.1185-1196(1990). [2]
```

```
RP SEQUENCE FROM N.A.
```

```
RC TISSUE=LIVER;
```

```
RX MEDLINE=87040757; PubMed=2430261;
```

```
RA Kaumeyer J.F., Pollazzi J.O., Kotick M.P.;
```

```
RT "The mRNA for a proteinase inhibitor related to the HI-30 domain of
```

```
RT inter-alpha-trypsin inhibitor also encodes alpha-1-microglobulin
```

```
RT (protein HC)." ;
```

```
RN Nucl. Acids Res. 14:7839-7850(1986). [3]
```

and causes the charge heterogeneity of protein HC.";
J. Biol. Chem. 266:15758-15763(1991).
[13] SEQUENCE OF 206-219, AND COVALENT LINKAGE WITH CHONDROITIN SULFATE.
TISSUE=PLASMA;
MEDLINE=94229087; PubMed=7513643;
Morelle W., Capon C., Balduyck M., Sautiere P., Kouach M.,
Michalski C., Fournet B., Mizon J.;
"Chondroitin sulphate covalently cross-links the three polypeptide
chains of inter-alpha-trypsin inhibitor.";
Eur. J. Biochem. 221:881-888(1994).
[14] SEQUENCE OF 206-223, AND CROSS-LINK SITE TO HC2.
MEDLINE=93232026; PubMed=7682553;
Engbild J.J., Salvesen G., Thøgersen I.B., Valnickova Z.,
Pizzo S.V., Hefta S.A.;
"Presence of the protein-glycosaminoglycan-protein covalent cross-link
in the inter-alpha-inhibitor-related proteinase inhibitor heavy chain
2/bikunin.";
J. Biol. Chem. 268:8711-8716(1993).
[15] SEQUENCE OF 206-223, AND CROSS-LINK SITE TO HC3.
MEDLINE=91093267; PubMed=1898736;
Engbild J.J., Salvesen G., Hefta S.A., Thøgersen I.B.,
Rutherford S., Pizzo S.V.;
"Chondroitin 4-sulfate covalently cross-links the chains of the human
blood protein pre-alpha-inhibitor.";
J. Biol. Chem. 266:747-751(1991).
[16] X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 230-339.
MEDLINE=98227321; PubMed=9566199;
Xu Y., Carr P.D., Guss J.M., Ollis D.L.;
"The crystal structure of bikunin from the inter-alpha-inhibitor
complex: a serine protease inhibitor with two Kunitz domains.";
J. Mol. Biol. 276:955-966(1998).
-1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
AND ALBUMIN.
-1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
URINE, INHIBITS TRYPsin, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
ELASTASE. ADDITIONAL PROTEOLYTIC PROCESSING IN THE KIDNEY AND/OR
URINE CAN PRODUCE FURTHER AMINO- AND CARBOXYL-END MODIFICATIONS
IN ITS SEQUENCE.
-1- SUBUNIT: INTER-ALPHA-TRYPsin INHIBITOR CONSIST OF A LIGHT CHAIN
AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.
-1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
SEPARATELY FUNCTIONING PROTEINS.
-1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.
-1- PTM: ADDITION OF GLYCOSAMINOGLYCAN CHONDROITIN SULFATE, ALLOWS
CROSS-LINKING BETWEEN THE DIFFERENT COMPONENTS.
-1- MISCELLANEOUS: IN VITRO, THE FIRST TWELVE RESIDUES OF THE AMINO
END OF THE INHIBITOR APPEAR TO HAVE A REACTIVE SITE CAPABLE OF
INHIBITING THE ACTIVITY OF A NUMBER OF ENZYMES. ITS IN VIVO
FUNCTION IS NOT KNOWN.
-1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
FAMILY.
-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE BPTI/KUNITZ
FAMILY OF INHIBITORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X54816; CAA38585.1;
DR EMBL; X54817; CAA38585.1; JOINED.
DR EMBL; X54818; CAA38585.1; JOINED.
DR EMBL; X04225; CAA27803.1;
DR EMBL; M88249; AAA59196.1;

SEQUENCE FROM N.A.
TISSUE=LIVER;
MEDLINE=90336621; PubMed=1696200;
Diarra-Mehrpour M., Bourguignon J., Sesboue R., Salier J.P.,
Leveillard T., Martin J.P.;
"Structural analysis of the human inter-alpha-trypsin inhibitor
"light-chain gene.";
Eur. J. Biochem. 191:131-139(1990).
[4] SEQUENCE OF 1-220 FROM N.A.
MEDLINE=86312901; PubMed=2428011;
Traboni C., Cortese R.;
"Sequence of a full length cDNA coding for human protein HC (alpha 1
microglobulin).";
Nucleic Acids Res. 14:6340-6340(1986).
[5] SEQUENCE OF 20-202 (INDIVIDUAL WITH TUBULAR PROTEINURIA).
MEDLINE=84126849; PubMed=6198962;
Lopez C., Grubb A.O., Mendez E.;
"The complete amino acid sequence of human complex-forming
glycoprotein heterogeneous in charge (protein HC) from one
individual.";
Arch. Biochem. Biophys. 228:544-554(1984).
[6] SEQUENCE OF 20-198 (VARIANT).
Lopez C., Grubb A.O., Mendez E.;
"Human protein HC displays variability in its carboxyl-terminal amino
acid sequence.";
FEBS Lett. 144:349-353(1982).
[7] SEQUENCE OF 20-198 (PATIENTS WITH TUBULAR PROTEINURIA).
MEDLINE=81184038; PubMed=6164372;
Takagi T., Takagi K., Kawai T.;
"Complete amino acid sequence of human alpha 1-microglobulin.";
Biochem. Biophys. Res. Commun. 98:997-1001(1981).
[8] SEQUENCE OF 206-350.
MEDLINE=85225968; PubMed=2408638;
Reisinger P., Hochstrasser K., Albrecht G.J., Lempart K., Salier J.P.;
"Human inter-alpha-trypsin inhibitor: localization of the Kunitz-type
domains in the N-terminal part of the molecule and their release by a
trypsin-like proteinase.";
Biol. Chem. Hoppe-Seyler 366:479-483(1985).
[9] CARBOHYDRATE-LINKAGE SITES.
MEDLINE=82074265; PubMed=6171497;
Hochstrasser K., Schoenberger O.L., Rossmann I., Wächter E.;
"Kunitz-type proteinase inhibitors derived by limited proteolysis of
the inter-alpha-trypsin inhibitor, V. Attachments of carbohydrates in
the human urinary trypsin inhibitor isolated by affinity
chromatography.";
Hoppe-Seyler's Z. Physiol. Chem. 362:1357-1362(1981).
[10] INHIBITORY SITE.
MEDLINE=85225940; PubMed=3890890;
Morli M., Travis J.;
"The reactive site of human inter-alpha-trypsin inhibitor is in the
amino-terminal half of the protein.";
Biol. Chem. Hoppe-Seyler 366:19-21(1985).
[11] STRUCTURE OF CARBOHYDRATES.
MEDLINE=90306345; PubMed=1694784;
Escibano J., Lopez-Otin C., Herpe A., Grubb A., Mendez E.;
"Location and characterization of the three carbohydrate prosthetic
groups of human protein HC.";
FEBS Lett. 266:167-170(1990).
[12] CHROMOPHORE.
TISSUE=URINE;
MEDLINE=91340714; PubMed=1714898;
Escibano J., Grubb A., Calero M., Mendez E.;
"The protein HC chromophore is linked to the cysteine residue at
position 34 of the polypeptide chain by a reduction-resistant bond

17 T_SITE 18 REACTIVE BOND (BY SIMILARITY).

Db 2 EDSQDLHAQPCGLGMISRYFYNGTSMACETFOYGGCLGNNGNFFASQKECLQTC 55

```
RESULT 8
IATR_SHEEP STANDARD; PRT; 123 AA.
AC PI3371;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1994 (Rel. 14, Last annotation update)
DE INTER-ALPHA-TRYPsin INHIBITOR (ITI) (GIK-14) (INHIBITORY FRAGMENT OF
DE ITI) (FRAGMENT).
OS Ovis aries (Sheep), and Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Caprinae; Ovis.
RN [1]
RP SEQUENCE.
RC SPECIES=SHEEP;
RX MEDLINE=87299012; PubMed=2441725;
RA Rasp G., Hochstrasser K., Wachter E., Reisinger P.W.M.;
RT "The amino-acid sequence of the trypsin-released inhibitor from sheep
inter-alpha-trypsin inhibitor".
RL Biol. Chem. Hoppe-Seyler 368:727-731(1987).
RN [2]
RP SEQUENCE.
RC SPECIES=C. HIRCUS;
RX MEDLINE=90105540; PubMed=2481505;
RA Rasp G., Hochstrasser K., Gerl C., Wachter E.;
RT "Primary structure of a proteinase inhibitor released from goat serum
inter-alpha-trypsin inhibitor".
RL Biochim. Biophys. Acta 999:335-337(1989).
CC -!- FUNCTION: THIS INHIBITORY FRAGMENT, RELEASED FROM NATIVE ITI AFTER
LIMITED PROTEOLYSIS WITH TRYPSIN, CONTAINS TWO HOMOLOGOUS DOMAINS.
WHEREAS THE SECOND DOMAIN IS A STRONG INHIBITOR OF TRYPSIN, THE
FIRST DOMAIN INTERACTS WEAKLY WITH PMN-GRANULOCYTIC ELASTASE AND
NOT AT ALL WITH PANCREATIC ELASTASE.
CC -!- MISCELLANEOUS: THE AMINO ACID AT POSITION P2' (17) APPEARS TO
DETERMINE THE SPECIFICITY OF THE INHIBITION OF DOMAIN I.
CC INHIBITORS WITH METHIONINE IN THIS POSITION INTERACT WEAKLY WITH
CHYMOTRYPSIN AND ELASTASE; THOSE WITH LEUCINE INTERACT STRONGLY.
CC -!- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
DR PIR: A29652; A29652.
DR HSSP: P10646; IADZ.
DR INTERPRO: IPR002223; -.
DR PFAM: PF00014; Kunitz_BPTI; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 2.
KW Plasma; Glycoprotein; Serine protease inhibitor; Repeat.
FT NON_TER 1 56 I.
FT DOMAIN 57 123 II.
FT DISULFID 5 55 BY SIMILARITY.
FT DISULFID 14 38 BY SIMILARITY.
FT DISULFID 30 51 BY SIMILARITY.
FT DISULFID 61 111 BY SIMILARITY.
FT DISULFID 70 94 BY SIMILARITY.
FT DISULFID 86 107 BY SIMILARITY.
FT ACT_SITE 15 16 INHIBITORY SITE (P1) (CHYMOTRYPSIN,
ELASTASE).
FT ACT_SITE 71 72 INHIBITORY SITE (P1) (TRYPSIN).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .).
FT NON_TER 123 123
SQ SEQUENCE 123 AA; 13686 MW; 295038173F2D2D2D1 CRC64;
```

Query Match 23.6%; Score 127; DB 1; Length 123;
Best Local Similarity 46.3%; Pred. No. 1.2e-07;
Matches 25; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 38 KDCKLDNMGSCYEVHFRYFNRTSKRCETFFVSGCNGNLFNFKLIEREVACVAKY 91

Db 2 EDSQDLGYSGQPCGLGMKFRFYNGTSMACETFOYGGCLGNNGNFFASQKECLQTC 55

```
RESULT 9
IBPC_BOVIN STANDARD; PRT; 67 AA.
AC P00976;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE COLOSTRUM TRYPSIN INHIBITOR (COLOSTRUM BPI).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE.
RA Cechova D., Jonakova V., Sorm F.;
RT "Primary structure of trypsin inhibitor from cow colostrum (component B2)".
RL Collect. Czech. Chem. Commun. 36:3342-3357(1971).
RN [2]
RP DISULFIDE BONDS.
RA Cechova D., Ber E.;
RT "Disulfide bonds of trypsin inhibitor from cow colostrum."
RL Collect. Czech. Chem. Commun. 39:680-688(1974).
RN [3]
RP CHARACTERIZATION.
RA Cechova D., Muszynska G.;
RT "Role of lysine 18 in active center of cow colostrum trypsin inhibitor."
RL FEBS Lett. 8:84-86(1970).
CC -!- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
DR PIR: A01207; TIBOC.
DR HSSP: P10646; ITEX.
DR INTERPRO: IPR002223; -.
DR PFAM: PF00014; Kunitz_BPTI; 1.
DR PRINTS: PR00759; BASICPTASE.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 1.
KW Serine protease inhibitor; Glycoprotein.
FT DISULFID 8 58
FT DISULFID 17 41
FT DISULFID 33 54 N-LINKED (GLCNAC. . .).
FT CARBOHYD 27 27 REACTIVE BOND (TRYPSIN).
FT ACT_SITE 18 19
FT ACT_SITE 67 67 E2B2093B7CD207CD CRC64;
SQ SEQUENCE 67 AA; 7511 MW; E2B2093B7CD207CD CRC64;
```

Query Match 23.4%; Score 126; DB 1; Length 67;
Best Local Similarity 41.0%; Pred. No. 8.4e-08;
Matches 25; Conservative 6; Mismatches 26; Indels 4; Gaps 1;

QY 39 DPCKLDNMGSCYEVHFRYFNRTSKRCETFFVSGCNGNLFNFKLIEREVACVAKYKPP 98

Db 6 DJCOLPOARGPCRAALLRYFNSTSNACEFTYGGCGGNBNF----ETTEMCLRICEPP 61

QY 99 R 99

Db 62 Q 62

```
RESULT 10
ELAC_TRIVU STANDARD; PRT; 102 AA.
AC Q29143;
ID ELAC_TRIVU
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE EARLY LACTATION PROTEIN PRECURSOR.
GN ELP.
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
```

RN SEQUENCE FROM N.A., AND SEQUENCE OF 21-30.
 RC TISSUE=MAMMARY GLAND;
 RA Plotte C.P., Grigor M.R.;
 RT "A novel marsupial protein expressed by the mammary gland only during
 RT the early lactation and related to the Kunitz proteinase inhibitors.";
 RL Arch. Biochem. Biophys. 330:59-64(1996).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: EXPRESSED BY THE MAMMARY GLAND.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY DURING EARLY LACTATION PHASE.
 CC -1- PTM: N-GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U34208; AAB08977.1; -
 DR HSP: P00974; 1PIT.
 DR INTERPRO: IPR002223; -
 DR PRINTS: PR00759; BASICPTASE.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_2; 1.
 KW Serine protease inhibitor; Signal; Lactation.
 FT SIGNAL 1 20
 FT CHAIN 21 102 EARLY LACTATION PROTEIN.
 FT DISULFID 43 93 BY SIMILARITY.
 FT DISULFID 52 76 BY SIMILARITY.
 FT DISULFID 68 89 BY SIMILARITY.
 FT ACT_SITE 53 54 REACTIVE BOND (BY SIMILARITY).
 FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 102 AA; 11407 MW; D769CA327D2E1BD3 CRC64;

Query Match 23.4%; Score 126; DB 1; Length 102;
 Best Local Similarity 56.1%; Pred. No. 1.3e-07;
 Matches 23; Conservative 3; Mismatches 15; Indels 0; Gaps 0;
 QY 41 CKLDNFGCVHFYFNRTSKRCFTVFSGNGNLNF 81
 DB 43 CLLPGRGNCDSQLRIFYNATSHTCVFLYSGNGNGNF 83

RESULT 11
 TFP2_HUMAN
 ID TFP2_HUMAN STANDARD; PRT; 235 AA.
 AC P48307;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE TISSUE FACTOR PATHWAY INHIBITOR 2 PRECURSOR (TFPI-2) (PLACENTAL
 DE PROTEIN 5) (P5).
 GN TFP12.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=PLACENTA;
 RX MEDLINE=95204397; PubMed=7896752;
 RA Miyagi Y., Koshikawa N., Yasumitsu H., Miyagi E., Hirahara F.,
 RA Aoki I., Misugi K., Umeda M., Miyazaki K.;
 RT "cDNA cloning and mRNA expression of a serine proteinase inhibitor
 RT secreted by cancer cells: identification as placental protein 5 and
 RT tissue factor pathway inhibitor-2.";
 RL J. Biochem. 116:939-942(1994).
 RN [2]

RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE=94211862; PubMed=8159751;
 RA Sprecher C.A., Kiesel W., Mathewes S., Foster D.C.;
 RT "Molecular cloning, expression, and partial characterization of a
 RT second human tissue-factor-pathway inhibitor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3353-3357(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Maggi L.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP PARTIAL SEQUENCE OF 23-35; 47-53 AND 133-146.
 RC TISSUE=PLACENTA;
 RX MEDLINE=88106628; PubMed=3276312;
 RA Buetzow R., Huhtala M.-L., Bohn H., Virtanen I., Seppaelae M.;
 RT "Purification and characterization of placental protein 5.";
 RL Biochem. Biophys. Res. Commun. 150:483-490(1988).
 RN [5]
 RP ERRATUM.
 RA Buetzow R., Huhtala M.-L., Bohn H., Virtanen I., Seppaelae M.;
 RL Biochem. Biophys. Res. Commun. 151:630-631(1988).
 CC -1- FUNCTION: SEEMS TO INHIBIT TRYPSIN, FACTOR VII(A)/TISSUE FACTOR,
 CC WEAKLY FACTOR XA. HAS NO EFFECT ON THROMBIN.
 CC -1- TISSUE SPECIFICITY: UMBILICAL VEIN ENDOTHELIAL CELLS, LIVER,
 CC PLACENTA, HEART, PANCREAS, AND MATERNAL SERUM AT ADVANCED
 CC PREGNANCY.
 CC -1- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
 CC HIGHLY SIMILAR TO TPFI.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D29992; BAA06272.1; -
 DR EMBL: L27624; AAA20094.1; -
 DR EMBL: AC002076; AAB54049.1; -
 DR PIR: A34029; A34029.
 DR PIR: B34029; B34029.
 DR PIR: C34029; C34029.
 DR HSP: P12111; IKNT.
 DR MIM: 600033; -
 DR INTERPRO: IPR002223; -
 DR PFAM: PF00014; Kunitz_BPTI; 3.
 DR PRINTS: PR00759; BASICPTASE.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
 DR PROSITE: PS00280; BPTI_KUNITZ_2; 3.
 KW Serine protease inhibitor; Glycoprotein; Repeat; Signal;
 KW Blood coagulation.
 FT SIGNAL 1 22
 FT CHAIN 23 235 TISSUE FACTOR PATHWAY INHIBITOR 2.
 FT DOMAIN 36 86 BPTI/KUNITZ INHIBITOR 1.
 FT DOMAIN 96 149 BPTI/KUNITZ INHIBITOR 2.
 FT DOMAIN 158 208 BPTI/KUNITZ INHIBITOR 3.
 FT DOMAIN 213 217 POLY-LYS.
 FT ACT_SITE 46 47 REACTIVE BOND (BY SIMILARITY).
 FT ACT_SITE 107 108 REACTIVE BOND (BY SIMILARITY).
 FT ACT_SITE 168 169 REACTIVE BOND (BY SIMILARITY).
 FT DISULFID 36 86 BY SIMILARITY.
 FT DISULFID 45 69 BY SIMILARITY.
 FT DISULFID 61 82 BY SIMILARITY.
 FT DISULFID 96 149 BY SIMILARITY.
 FT DISULFID 106 130 BY SIMILARITY.
 FT DISULFID 122 145 BY SIMILARITY.
 FT DISULFID 158 208 BY SIMILARITY.
 FT DISULFID 167 191 BY SIMILARITY.
 FT DISULFID 183 204 BY SIMILARITY.
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT CARBOHYD 170 170 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 23 D -> A (IN REF. 4).
SQ SEQUENCE 235 AA; 26934 MW; 975ABA5C53F7C65F CRC64;

Query Match 23.0%; Score 124; DB 1; Length 235;
Best Local Similarity 31.4%; Pred. No. 5.2e-07;
Matches 27; Conservative 18; Mismatches 37; Indels 4; Gaps 1;

QY 18 LNTLL---GGVNKIAEKIGDLKDPCKLDMNFGSCYEVHFRYFRTSKRCETFFVSG 73
   ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
   9 LSIILLFLTEALGDAQAEFTGNNAEICLLPLDYGPCRALLRYYDRYQSCROFLYGG 68
   ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 74 CNGNLFNFKLKIEREVACVAKYKPPR 99
   ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 69 CEGNANNFYTWACDADACWRKVPK 94
   ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

RESULT 12
AMP_MOUSE STANDARD; PRT; 349 AA.
AC Q07456; 061294;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AMP PROTEIN PRECURSOR (CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)).
GN AMP OR ITIL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=LIVER;
RX MEDLINE=93363639; PubMed=7689339;
RA Chan P., Salier J.P.;
RT "Mouse alpha-1-microglobulin/bikunin precursor: cDNA analysis, gene
RT evolution and physical assignment of the gene next to the orosomucoid
RT locus.";
RL Biochim. Biophys. Acta 1174:195-200(1993).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RX MEDLINE=95189774; PubMed=7533761;
RA Itoh H., Ide H., Kataoka H., Tomita M., Yoshihara H., Nawa Y.;
RT "cDNA sequencing of mouse alpha 1-microglobulin/inter-alpha-trypsin
RT inhibitor light chain and its expression in acute inflammation.";
RL J. Biochem. 116:767-772(1994).
[3]
RN SEQUENCE OF 128-349 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RA Itoh H., Ide H., Yoshihara H., Nawa Y.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC AND ALBUMIN (BY SIMILARITY).
CC !- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYOSOMAL GRANULOCYTIC
CC ELASTASE (BY SIMILARITY).
CC !- SUBUNIT: INTER-ALPHA-TRYPSIN INHIBITOR CONSIST OF A LIGHT CHAIN
CC AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.
CC !- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC SEPARATELY FUNCTIONING PROTEINS.
CC !- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
CC SIMILARITY).
CC !- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
CC FAMILY.
CC !- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE BPTI/KUNITZ
CC FAMILY OF INHIBITORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```


Search completed: March 24, 2001, 13:24:21
Job time: 167 sec

QY		91 C 91
Db		280 C 280
RESULT	15	
TFPI_RABIT		
ID	TEPI_RABIT	STANDARD; PRT; 300 AA.
AC	R19761; Q28828;	
DT	01-FEB-1991 (Rel. 17, Created)	
DT	01-AUG-1992 (Rel. 23, Last sequence update)	
DT	01-NOV-1997 (Rel. 35, Last annotation update)	
DE	TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-	
DE	ASSOCIATED COAGULATION INHIBITOR) (LACI) (EXTRINSIC PATHWAY INHIBITOR)	
DE	(EPI).	
GN	TFPI.	
OS	Oryctolagus cuniculus (Rabbit).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.	
[1]		
RP	SEQUENCE FROM N.A.	
RC	TISSUE=LIVER:	
RA	MEDLINE=91057146; PubMed=2136251;	
RX	Wesselschmidt R.L., Girard T.J., Broze G.J. Jr.;	
RT	"cDNA sequence of rabbit lipoprotein-associated coagulation	
RT	inhibitor."	
RL	Nucleic Acids Res. 18:6440-6440(1990).	
[2]		
RP	REVISIONS TO 72; 211 AND 218.	
RC	TISSUE=LIVER:	
RC	MEDLINE=92335027; PubMed=1630940;	
RX	Warn-Cramer B.J., Broze G.J. Jr., Komives E.A.;	
RT	"cDNA sequence of rabbit tissue factor pathway inhibitor.";	
RL	Nucleic Acids Res. 20:3548-3548(1992).	
[3]		
RP	SEQUENCE FROM N.A.	
RC	TISSUE=LUNG:	
RC	MEDLINE=93276427; PubMed=8503123;	
RX	Belacouaj A., Kuppuswamy M.N., Birktoft J.J., Bajaj S.P.;	
RA	"Revised cDNA sequence of rabbit tissue factor pathway inhibitor.";	
RT	Thromb. Res. 69:547-553(1993).	
CC	-!- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT	
CC	WAY, INHIBITS VII(A)/TISSUE FACTOR ACTIVITY, PRESUMABLY BY FORMING	
CC	A QUATERNARY X(A)/LACI/VII(A)/TF COMPLEX. IT POSSESSES AN	
CC	ANTI-THROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH	
CC	LIPOPROTEINS IN PLASMA.	
CC	-!- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.	
CC	-!- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.	
CC	HIGHLY SIMILAR TO TFP2.	
CC		
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-	
CC	tion between the Swiss Institute of Bioinformatics and the EMBL Outstat-	
CC	the European Bioinformatics Institute. There are no restrictions on	
CC	use by non-profit institutions as long as its content is in no	
CC	modified and this statement is not removed. Usage by and for commer-	
CC	entities requires a license agreement (See http://www.isb-sib.ch/anno-	
CC	or send an email to licensese@isb-sib.ch).	
CC		
DR	EMBL; X54708; CAA38515.1; ALT_SEQ.	
DR	EMBL; S61902; AAB28836.1; -	
DR	PIR; S12143; S12143.	
DR	HSSP; P10646; ITPF.	
DR	INTERPRO; IPRO02223; -	
DR	PFAM; PF00014; Kunitz-BPTI; 3.	
DR	PRINTS; PR00759; BASICPTASE.	
DR	PROSITE; PS00280; BPTI_KUNITZ_1; 3.	
DR	PROSITE; PS02079; BPTI_KUNITZ_2; 3.	
KW	Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation;	
KW	Signal.	
FT	1 24	
FT	CHAIN 25 300	TISSUE FACTOR PATHWAY INHIBITOR.
FT	DOMAIN 50 100	BPTI/KUNITZ INHIBITOR 1

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:22:25 ; Search time 47.63 Seconds
(without alignments)
71.073 Million cell updates/sec

Title: US-09-215-435-177

Perfect score: 539

Sequence: 1 MKSAKLGFLRRFFIFCSLNT.....NFKLKIERVACVAKYKPPR 99

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	539	100.0	99	20 Y35928	Extended human sec
2	529	98.1	99	20 Y35054	Extended human sec
3	427	79.2	80	20 Y12791	Human 5' EST secre
4	142	26.3	122	20 Y08611	Inter-alpha-trypsi
5	140	26.0	144	18 W25938	SUC2-Ep7-G21-RPDP-
6	138	25.6	144	18 W25936	SUC2-Ep1-G21-RPDP-
7	135	25.0	58	16 R81936	Human ITI-K1 Kunit
8	133.5	24.8	560	15 R62523	Hookworm anticoagu
9	132	24.5	58	16 R78542	Human LACI-K2 deri
10	131.5	24.4	165	18 R78537	Elastase inhibitor
11	131	24.3	58	16 R78554	Human ITI-KuDOM 1
12	131	24.3	95	17 R99213	ITI-D1 Kunitz doma

13	130	24.1	58	17 R99191	Genetically engine
14	129.5	24.0	143	20 Y08609	Inter-alpha-trypsi
15	129.5	24.0	147	17 R92237	Human wild-type ur
16	129.5	24.0	147	18 W25928	Anti-trypsin inhib
17	129.5	24.0	147	19 W69522	ruti protein seq 1
18	129.5	24.0	352	9 P81110	Sequence of new fu
19	129	23.9	58	17 R99196	Genetically engine
20	128	23.7	58	13 R27402	Human neutrophil e
21	128	23.7	58	17 R99160	Genetically engine
22	127.5	23.7	145	17 R92235	Human urinary tryp
23	127	23.6	55	17 R99217	Inter alpha trypsi
24	127	23.6	58	16 R78553	Human ITI Kunitz d
25	127	23.6	58	17 R99195	Genetically engine
26	127	23.6	58	17 R99197	Genetically engine
27	127	23.6	58	19 W64115	Human Kunitz-type
28	127	23.6	58	20 W92862	US5880256 Seq ID 3
29	127	23.6	111	14 R39807	Synthetic yeast ie
30	126	23.4	58	13 R27394	Human neutrophil e
31	126	23.4	58	17 R99157	Human aprotinin-11
32	126	23.4	58	17 R99161	Genetically engine
33	126	23.4	58	17 R99162	Genetically engine
34	126	23.4	124	18 W25933	New protease inhib
35	125	23.2	58	17 R99158	Genetically engine
36	125	23.2	58	17 R99193	Genetically engine
37	125	23.2	58	17 R99164	Genetically engine
38	125	23.2	60	11 R08291	Human bikunin doma
39	125	23.2	74	17 R92238	Human UTI modified
40	125	23.2	124	17 R92233	Human mature urina
41	124	23.0	55	18 W25931	RPDP-Kunitz domain
42	124	23.0	58	17 R99194	Genetically engine
43	124	23.0	124	18 W25932	New protease inhib
44	124	23.0	145	18 W25935	Novel protease inh
45	124	23.0	235	16 R74977	Human Kunitz-type

ALIGNMENTS

RESULT 1

Y35928
ID Y35928 standard; Protein; 99 AA.
XX
XX Y35928;
AC
XX
DF 13-SEP-1999 (first entry)
XX
DE Extended human secreted protein sequence, SEQ ID NO. 177.

Secreted protein; human; cytokine; cellular proliferation; cell movement;
cellular differentiation; immune system regulator; anti-inflammatory;
haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
genetic disease.

OS Homo sapiens.

XX

PN WO9931236-A2.

XX

PD 24-JUN-1999.

XX

PF 17-DEC-1998; 98WO-IB02122.

XX

PR 10-AUG-1998; 98US-0096116.

PR 17-DEC-1997; 97US-0069957.

PR 09-FEB-1998; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

XX (GEST) GENSET.

PA

XX Bougueleret L, Duclert A, Dumas Milne Edwards J;

PI WPI; 1999-385906/32.

DR N-PSDB; X97612.

DR

XX PT New isolated human secreted proteins
XX PS Claim 9; Page 210; 516pp; English.

XX CC This sequence is encoded by an extended human secreted protein coding
XX CC sequence of the invention. The secreted proteins can be used in treating
XX CC or controlling a variety of human conditions. The secreted proteins may
XX CC act as cytokines or may affect cellular proliferation or differentiation
XX CC or may act as immune system regulators, haematopoiesis regulators, tissue
XX CC growth regulators, regulators of reproductive hormones or cell movement
XX CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
XX CC tumour inhibition activity. The DNAs can be used in forensic procedures
XX CC to identify individuals or in diagnostic procedures to identify
XX CC individuals having genetic diseases resulting from abnormal expression of
XX CC the genes corresponding to the extended cDNAs. They are also useful for
XX CC constructing a high resolution map of the human chromosomes. They can
XX CC also be used for gene therapy to control or treat genetic diseases.

XX SQ Sequence 99 AA;

Query Match 100.0%; Score 539; DB 20; Length 99;
Best Local Similarity 100.0%; Pred. No. 4e-57;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSAKLGFLRRFFIFCSLNTLLGGVKNIAEKICGDLKDPCKLDMNFGSCYEVHFRFYFN 60
Db 1 mksaklgflrrffifcsntlllggvnkieakicgdlkdpckldmnmfgscyevhfrfyfn 60
Qy 61 RTSKRCETFFVSCNGNLNFKLIEREVACVAKYKPPR 99
Db 61 rtskrctfvfscngnlnnflklriervacvakykppr 99

RESULT 2

Y36054 ID Y36054 standard; Protein; 99 AA.

XX AC Y36054;

XX DT 13-SEP-1999 (first entry)

XX DE Extended human secreted protein sequence, SEQ ID NO. 439.

XX KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
XX KW cellular differentiation; immune system regulator; anti-inflammatory;
XX KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
XX KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
XX KW genetic disease.

XX OS Homo sapiens.

XX PN WO9931236-A2.

XX PD 24-JUN-1999.

XX PF 17-DEC-1998; 98WO-IB02122.

XX PR 10-AUG-1998; 98US-0096116.

XX PR 17-DEC-1997; 97US-0069957.

XX PR 09-FEB-1998; 98US-0074121.

XX PR 13-APR-1998; 98US-0081563.

XX PA (GEST) GENSET.

XX PI Bouqueleret L, Duclet A, Dumas Milne Edwards J;

XX DR WPI; 1999-385906/32.

XX DR N-PSDB; X97738.

XX PT New isolated human secreted proteins

PS Claim 9; Page 376; 516pp; English.

XX CC This sequence is encoded by an extended human secreted protein coding
XX CC sequence of the invention. The secreted proteins can be used in treating
XX CC or controlling a variety of human conditions. The secreted proteins may
XX CC act as cytokines or may affect cellular proliferation or differentiation
XX CC or may act as immune system regulators, haematopoiesis regulators, tissue
XX CC growth regulators, regulators of reproductive hormones or cell movement
XX CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
XX CC tumour inhibition activity. The DNAs can be used in forensic procedures
XX CC to identify individuals or in diagnostic procedures to identify
XX CC individuals having genetic diseases resulting from abnormal expression of
XX CC the genes corresponding to the extended cDNAs. They are also useful for
XX CC constructing a high resolution map of the human chromosomes. They can
XX CC also be used for gene therapy to control or treat genetic diseases.

XX SQ Sequence 99 AA;

Query Match 98.1%; Score 529; DB 20; Length 99;
Best Local Similarity 98.0%; Pred. No. 6.3e-56;
Matches 97; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSAKLGFLRRFFIFCSLNTLLGGVKNIAEKICGDLKDPCKLDMNFGSCYEVHFRFYFN 60
Db 1 mksaklgflrrffifcsntlllggvnkieakicgdlkdpckldmnmfgscyevhfrfyfn 60
Qy 61 RTSKRCETFFVSCNGNLNFKLIEREVACVAKYKPPR 99
Db 61 rtskrctfvfscngnlnnflklriervacvakykppr 99

RESULT 3

Y12791 ID Y12791 standard; Protein; 80 AA.

XX AC Y12791;

XX DT 21-JUN-1999 (first entry)

XX DE Human 5' EST secreted protein SEQ ID NO:381.

XX KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX KW forensic; gene therapy; chromosome mapping; signal peptide;
XX KW upstream regulatory sequence; cytokine activity; cell proliferation;
XX KW differentiation; haematopoiesis regulation; tissue growth regulation;
XX KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX KW thrombolytic; anti-inflammatory; tumour inhibition.

XX OS Homo sapiens.

XX PN WO9906549-A2.

XX PD 11-FEB-1999.

XX PF 31-JUL-1998; 98WO-IB01231.

XX PR 01-AUG-1997; 97US-0905279.

XX PA (GEST) GENSET.

XX PI Duclert A, Dumas Milne Edwards J, Lacroix B;

XX PI WPI; 1999-153779/13.

XX DR N-PSDB; X51569.

XX PT New nucleic acids encoding human secreted proteins - obtained from
XX PT cDNA libraries derived from testis, ovary, uterus and spleen tissue

XX PS Claim 34; Page 449; 522pp; English.

XX CC X51459 to X51691 represent 5' expressed sequence tags (ESTs) for human
XX CC secreted proteins, and encode the proteins given in Y12681 to Y12913,

07-NOV-1995;	95JP-0288527.
07-NOV-1995;	95JP-0288527.
(GRC)	GREEN CROSS CORP.
WPI;	1997-316576/29.
N-PSDB;	T79038.
New protease inhibitor - useful for treating diseases involving elastase	
Disclosure; Fig 38; 37pp; Japanese.	
This is the amino acid sequence of the polypeptide encoded by the insert in plasmid PHH341 which comprises the novel elastase specific inhibitor Ep7-d21-RPDF-52-55 (W25933). The inhibitor sequence is linked downstream of the yeast invertase (SUC2) signal peptide sequence. The modified protease inhibitors are targeted to the protease elastase, especially from neutrophils and can be used to treat diseases associated with elastase. Modifications of the active site were done by site directed mutagenesis.	
Sequence	144 AA;
Query Match	26.0%; Score 140; DB 18; Length 144;
Best Local Similarity	41.0%; Pred. No. 2.5e-09;
Matches	34; Conservative 10; Mismatches 29; Indels 10; Gaps 2;
y	9 LLREFFCSLNTLLGGVKNKIAKICGLDKDPCKLDMFGSCYEVHFRVYVRTSKRCET 68
b	3 llgafif-----llagfaakisar-----pdfcqlgysagpcvamfryfyngtmacqt 52
y	69 FVFGSGCNGNLNFKLKIEREVAC 91
b	53 fvyggcmgnfnftekeclqtc 75
RESULT	6
D	225936
D	W25936 standard; Protein; 144 AA.
X	
X	W25936;
C	
12-NOV-1997	(first entry)
SUC2-Ep1-d21-RPDF-52-55	from plasmid pHH337.
Trypsin inhibitor;	Kunitz domain; protease; active site; elastase;
neutrophil; disease;	modification; site directed mutagenesis.
Synthetic.	
Key	Location/Qualifiers
Peptide	1..20
Protein	/note= "signal peptide"
Domain	21..144
Disulfide-bond	/note= "mature protein"
Disulfide-bond	21..76
Misc-difference	/note= "Kunitz domain 1"
Disulfide-bond	25..75
Disulfide-bond	34..58
Misc-difference	35..39
Disulfide-bond	/note= "mutated region; amino acids substituted for wild type amino acid sequence WGMPS; this includes the active site residue of domain 1 (aa 36 of the wild type sequence W25928)"
Disulfide-bond	50..71
Misc-difference	51
Disulfide-bond	/note= "mutated amino acid; replaces amino acid Glu in wild type sequence"

	Synthetic.	
XX		
OS		
XX		
XX		
Key	Location/Qualifiers	
Misc-difference	10	
Ft	/note=	"residue change: Ser to Glu"
Ft		
Ft	Misc-difference	15
Ft	/note=	"residue change: Met to Arg"
Ft		
Ft	Misc-difference	17
Ft	/note=	"residue change: Met to Lys"
Ft		
Ft	Misc-difference	18
Ft	/note=	"residue change: Thr to Phe"
Ft		
Ft	Misc-difference	34
Ft	/note=	"residue change: Gln to Val"
Ft		
Ft	Misc-difference	39
Ft	/note=	"residue change: Met to Gly"
Ft		

Key	Location/Qualifiers
Peptide	1..20
Protein	/note= "signal peptide"
Modified-site	21..165
Domain	/note= "mature protein"
Misc-difference	30
Disulfide-bond	/label= O-linked glycosylation site
Disulfide-bond	42..97
Disulfide-bond	/note= "Kunitz domain 1"
Disulfide-bond	42..45
Disulfide-bond	/note= "mutated amino acids: replaces wild type sequence KEDS"
Disulfide-bond	46..96
Disulfide-bond	55..79
Disulfide-bond	56..60
Disulfide-bond	/note= "mutated region; amino acids substituted for wild type amino acid sequence MGMTS; this includes the active site residue of domain 1"
Disulfide-bond	71..92
Disulfide-bond	72
Disulfide-bond	/note= "mutated amino acid; replaces amino acid Glu in wild type sequence"
Disulfide-bond	75
Disulfide-bond	/note= "mutated amino acid; replaces amino acid Gln in wild type sequence"
Disulfide-bond	98..165
Disulfide-bond	/note= "Kunitz domain 2"
Disulfide-bond	102..152
Disulfide-bond	111..135
Disulfide-bond	112
Disulfide-bond	/note= "active site residue"
Disulfide-bond	127..148

DR WPI: 1999-357231/30.

XX Transgenic mice useful for studying compounds potentially useful in
PT the treatment of Alzheimer's disease

XX

XX Disclosure; Fig 8A; 72pp; English.

PS

XX This invention describes novel transgenic mice expressing proteins
XX related to the pathology of Alzheimer's disease and which provide models
CC for studying potentially therapeutic compounds. The transgenic mice
CC contain a DNA sequence encoding a beta-amyloid precursor protein (APP)
CC and a nerve tissue specific promoter operably linked to the beta-APP
CC allowing its expression to form beta-amyloid protein deposits in the
CC animal's brain. The transgenic mouse is useful for elucidating the
CC molecular mechanisms involved in the synthesis of and, more importantly,
CC inhibiting the synthesis and deposition of beta-amyloid proteins (most
CC importantly in the brain where plaque formation is associated with
CC Alzheimer's disease) by inhibiting production and/or increasing cleavage
CC after production. The transgenic animals provide useful models for
CC studying the in vivo relationships of the proteins to each other and to
CC other compounds being tested for their usefulness in treating Alzheimer's
CC disease.

XX

XX Sequence 143 AA;

SO

Query Match 24.0%; Score 129.5; DB 20; Length 143;
Best Local Similarity 41.4%; Pred. No. 4.4e-08;
Matches 29; Conservative 25; Indels 7; Gaps 1

Qy 82 KLKIEREVAC 91
 ; |
ph 67 vrekeclatc 76
Db 12 ggqlvtevtkk-----edscqlgysagpcmgmtsrifnyngtmsacetrqygycmgungunni

ID	R92237 standard; protein; 147 AA.
XX	
AC	R92237;
XX	
DT	27-SEP-1996 (first entry)
XX	
DE	Human wild-type urinary trypsin inhibitor.
XX	
XX	UTI; kunitz domain; urinary trypsin inhibitor; elastase inhibitor;
KW	recombinant protein production; yeast host cell; Pichia;
KW	site-directed mutagenesis.

FT	peptide	1..21	/label= signal_peptide
FT	FT	26..76	
FT	Domain		/label= kunitz_domain_1
FT	FT		
FT	Misc-difference	36..40	/label= p1-p4'
FT	FT		/note= "motif replaced to improve elias"
FT	FT		
FT	Domain	78..145	/label= kunitz_domain_2
FT	FT		
FT	Disulfide-bond	26..76	
FT	Disulfide-bond	35..59	
FT	Disulfide-bond	51..72	
FT	Disulfide-bond	82..132	
FT	Disulfide-bond	91..115	
FT	Disulfide-bond	107..128	
FT	Modified-site	10	
FT	FT		

THIS PAGE BLANK (USPTO)

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	147.5	27.4	58	5	Q9TWF9	Q9TWF9 anemonia su	
2	146.5	27.2	58	5	Q9TWF0	Q9TWF0 anemonia su	
3	141	26.2	3198	5	Q9U868	Q9U868 manduca sex	
4	132.5	24.6	132	5	Q9VQT9	Q9VQT9 drosophila	
5	129.5	24.0	151	4	Q78491	Q78491 homo sapien	
6	128	23.7	2167	5	Q76840	Q76840 caenorhabdi	
7	126	23.4	59	5	Q9TWF8	Q9TWF8 anemonia su	
8	125.5	23.3	230	11	Q35536	Q35536 mus musculus	
9	124.5	23.1	249	5	Q16701	Q16701 caenorhabdi	
10	123.5	22.9	154	6	Q9NOX3	Q9NOX3 ovis aries	
11	123.5	22.9	2225	5	Q45881	Q45881 caenorhabdi	
12	122	22.6	287	13	Q93424	Q93424 cyprinus ca	
13	120	22.3	342	13	Q70004	Q70004 xenopus lae	
14	119.5	22.2	169	6	Q9NOX7	Q9NOX7 bos taurus	
15	119	22.1	246	11	Q92ZU8	Q92ZU8 mus musculus	
16	119	22.1	306	11	Q54819	Q54819 mus musculus	
17	118.5	22.0	396	6	Q28874	Q28874 canis famli	
18	117	21.7	805	5	Q19305	Q19305 caenorhabdi	
19	116	21.5	251	4	Q95103	Q95103 homo sapien	

5

DT	01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DE	01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE	HYPOTHEtical 33.1 KDA PROTEIN.
OS	Cyprinus carpio (Common carp).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC	Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
OX	NCBI_TaxID=7962;
RN	[1]
RC	SEQUENCE FROM N.A.
RA	TISSUE=LIVER;
RA	Gracey A.;
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF008648; AAC19410.1; -
DR	HSP; P31713; 1SHP.
DR	INTERPRO; IPR002223; -
DR	PRAM; PF00014; Kunitz_BPTI; 3.
DR	PRINTS; PR00759; BASICPTASE
DR	PROSITE; PS00280; BPTI_KUNITZ; 3.
KW	Hypothetical protein; Serine protease inhibitor.
SQ	SEQUENCE 287 AA; 33093 MW; DF69B3D76718115E CRC64;

Query Match	22.68;	Score 122;	DB 13;	Length 287;
Best Local Similarity	40.68;	Pred. No. 7.6e-07;		
Matches 26;	Conservative 10;	Mismatches 26;	Indels 2;	Gaps

QY	31	EKIC- -GDLDPCKLDMFGSCYEVHRYFNRTSKRCETVFSGCGNLNNFKLKIERE 88
Db	89	EKMCLVREDKSPQQLDDPQPCRGVLVPRYFDFKSEQCKRFFYGGCGFNANFKTKECH 148
QY	89	VACV 92
Db	149	ERCL 152

RESULT 13			
P70004	PRELIMINARY;	PRT;	342 AA.
ID	P70004		
AC	P70004;		
DT	01-FEB-1997 (TReMBLrel. 02, Created)		
DT	01-FEB-1997 (TReMBLrel. 02, Last sequence update)		
DE	01-JUN-2000 (TReMBLrel. 14, Last annotation update)		
DE	ALPHA1-MICROGLOBULIN/BIKUNIN PRECURSOR (AMBP).		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8355;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RA	TISSUE=LIVER;		
RA	Kawahara A., Hikosaka A., Sasado T., Hirota K.;		
RT	"Thyroid hormone-dependent repression of alpha1-microglobulin/bikunin		
RT	(AMBP) gene expression during amphibian metamorphosis.";		
RL	Dev. Genes Evol. 206:355-362(1997).		
DR	EMBL; D87752; BAA13453.1; -		
DR	HSP; P02760; IBIK.		
DR	INTERPRO; IPR000566; -		
DR	INTERPRO; IPR00223; -		
DR	INTERPRO; IPR002345; -		
DR	INTERPRO; IPR002968; -		
DR	PFAM; PF00014; Kunitz_BPTI; 2.		
DR	PFAM; PF00061; lipocalin; 1.		
DR	PRINTS; PR00179; LIPOCALIN.		
DR	PRINTS; PR00759; BASICPTASE.		
DR	PRINTS; PR01215; ALMGLOBULIN.		
DR	PROSITE; PS00213; LIPOCALIN; 1.		
DR	PROSITE; PS00280; BPTI_KUNITZ; 2.		
KW	Signal; Serine protease inhibitor.		
FT	SIGNAL 1 17	POTENTIAL.	
FT	CHAIN 18 342	POTENTIAL.	
SQ	SEQUENCE 342 AA; 38558 MW; 3050508BFDG5C47B CRC64;		

Query Match 22.3%; Score 120; DB 13; Length 342;
Best Local Similarity 42.4%; Pred. No. 1.6e-06;
Matches 25; Conservative 6; Mismatches 22; Indels 6; Gaps 1;
Qy 39 DCKLDMNFGSCYEVHFRFYNTSKRCETVFVSGNGNLNFKLIE-----REVAC 91
Db 223 ESCR LAPSGPC LIGNHRYFNYSNMACEFTQGGCLGNHNFHSEKECLHDCRTEAC 281
RESULT 14
Q9NOX7 PRELIMINARY; PRT; 169 AA.
AC Q9NOX7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE TROPHOBLAST KUNITZ DOMAIN PROTEIN 2.
GN TKDPI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BREED SIMMENTAL X HEREFORD;
RA Green J.A., Maclean J.A. II, Gan X., Xie S., Roberts R.M.;
RT "Cloning and characterization of the expression of trophoblast Kunitz
domain proteins (TKDP).";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF241777; AAF61248.1; -
SQ SEQUENCE 169 AA; 18855 MW; B65C16F4DF0BA3D CRC64;

Query Match 22.2%; Score 119.5; DB 6; Length 169;
Best Local Similarity 42.9%; Pred. No. 8.6e-07;
Matches 24; Conservative 5; Mismatches 18; Indels 9; Gaps 1;
Qy 36 DLKDPCKLDMNFGSCYEVHFRFYNTSKRCETVFVSGNGNLNFKLIEREVAC 91
Db 116 ELKGPCKDQMT-----RYFNATRYCEPFVYGGCGGNKNFQTLSHCIIVTC 162
RESULT 15
Q922U8 PRELIMINARY; PRT; 246 AA.
AC Q922U8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE TFP1BETA (FRAGMENT).
GN TFP1BETA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Chang J.-Y., Monroe D.M., Roberts H.R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF016313; AAD01586.1; -
DR HSP; P10646; ITEX
DR INTERPRO: IPR002223; -
DR PFAM: PF00014; Kunitz_BPTI; 2.
DR PRINTS: PR00759; BASICPTASE.
DR PROSITE: PS00280; BPTI_KUNITZ; 2.
KW Serine protease inhibitor.
FT NON_TER 1
SQ SEQUENCE 246 AA; 28003 MW; 4BEDC23EE04F07E6 CRC64;

Query Match 22.1%; Score 119; DB 11; Length 246;
Best Local Similarity 39.1%; Pred. No. 1.5e-06;
Matches 27; Conservative 8; Mismatches 32; Indels 2; Gaps 1;
Qy 25 GVNKIAEKICGDLKDP--CKLDMNFGSCYEVHFRFYNTSKRCETVFVSGNGNLNFK 82
Db 96 GYEXTAVKAASGAERPDFCFLEEDPGLCRGYMKRYLYNNOTKQERFVYGGCLGNRNFE 155
Qy 83 LKIEREVAC 91
Db 156 TLDECKKTC 164
Search completed: March 24, 2001, 13:25:41
Job time: 247 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:23:12 ; Search time 44.26 Seconds
(without alignments)
151.879 Million cell updates/sec

Title: US-09-215-435-177
Perfect score: 539
Sequence: 1 MSAKLGFLRRFFIFCSLNT.....NFKLKIEREVACVAKYKPPR 99

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_66:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154	28.6	110	1 TITROR	basic proteinase i
2	134	24.9	352	1 TIBOBI	alpha-1-microglobu
3	133	24.7	1558	2 T34394	hypothetical prote
4	130	24.1	62	2 S07451	proteinase inhibit
5	129.5	24.0	352	1 HCHU	alpha-1-microglobu
6	128	23.7	2167	2 T34395	hypothetical prote
7	127	23.6	65	1 TIVIVC	venom basic protei
8	127	23.6	123	2 A29652	inter-alpha-trypsi
9	127	23.6	125	1 TIBOBI	alpha-1-microglobu
10	126	23.4	67	1 TIBOC	trypsin inhibitor,
11	126	23.4	102	2 S69288	early lactation pr
12	124.5	23.1	249	2 T32060	hypothetical prote
13	124	23.0	235	2 A54951	tissue factor path
14	124	23.0	349	2 S35708	alpha-1-microglobu
15	123.5	22.9	2225	2 T26063	hypothetical prote
16	122	22.6	60	1 TIEPVI	venom basic protei
17	122	22.6	349	2 S21089	alpha-1-microglobu
18	121	22.4	299	2 I46937	tissue factor path
19	121	22.4	300	2 S12143	lipoprotein-associ
20	120	22.3	304	1 JC2264	tissue factor path
21	118.5	22.0	396	2 S53325	tissue factor path
22	118	21.9	57	2 A59204	basic proteinase i
23	118	21.9	62	2 S19327	venom basic protei
24	118	21.9	337	1 TIFGBI	alpha-1-microglobu
25	117	21.7	57	1 TIFHBP	proteinase inhibit
26	117	21.7	302	1 TIRTKG	tissue factor path
27	117	21.7	805	2 T34212	hypothetical prote
28	116	21.5	304	1 TIHUGK	tissue factor path
29	116	21.5	1743	2 T26859	hypothetical prote

30	115.5	21.4	1965	2 T33216	hypothetical prote
31	115	21.3	59	1 VIEPIA	venom basic protei
32	114	21.2	62	2 A4180	taicatoxin serine
33	113	21.0	1599	2 T16210	hypothetical prote
34	112.5	20.9	355	1 S22181	gamma-1-microglobu
35	112	20.8	1043	2 T19734	hypothetical prote
36	111	20.6	57	2 S12957	venom animal Kunit
37	111	20.6	372	2 JC2556	alpha-1-microglobu
38	110	20.4	252	2 JG0185	hepatocyte growth
39	110	20.4	922	2 T23573	hypothetical prote
40	109	20.2	60	2 A36989	calciclutide - eas
41	109	20.2	2944	2 A54849	collagen alpha 1(V
42	108.5	20.1	765	2 S42880	anlyoid precursor-
43	108	20.0	62	2 S01803	chymotrypsin inhib
44	107.5	19.9	751	2 A49574	beta-amyloid precu
45	107.5	19.9	763	2 A49321	amyloid beta (A4)

ALIGNMENTS

RESULT 1

TITROR

basic proteinase inhibitor - loggerhead

C:Species: Caretta caretta (loggerhead)

C:Date: 31-May-1979 #sequence_revision 23-Oct-1981 #text_change 05-Aug-1994

C:Accession: A01224

R:Kato, I.; Tomimaga, N.

Fed. Proc. 38, 832, 1979

A:Title: Trypsin--subtilisin inhibitor from red sea turtle eggwhite consists of two t

A:Reference number: A01224

A:Accession: A01224

A:Molecule type: protein

A:Residues: 1-110 <RAT>

C:Comment: This inhibitor, isolated from egg white, consists of two nonhomologous dom

C:Superfamily: loggerhead basic proteinase inhibitor; animal Kunitz-type proteinase i

C:Keywords: pyrrolutamic acid; serine proteinase inhibitor

F:8-58/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

F:63-105/Domain: antileukoproteinase repeat homology <ALP>

F:/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:8-58,17-41,33-54,76-97,80-92,86-101/Disulfide bonds: #status predicted

F:18/Inhibitory site: Lys (trypsin) #status predicted

Query Match 28.6%; Score 154; DB 1; Length 110;

Best Local Similarity 46.9%; Pred. No. 5.4e-10;

Matches 30; Conservative 7; Mismatches 23; Indels 4; Gaps 1;

QY 35 GDLKDPCKLDNFGSCYEVHFRFYFNRTSKRCETVFSGCNGNLNFKLKTEREVACVAK 94

Db 2 GDRKDICRLPPEQGPCGRIPRYFNPNASRCESFIYGGCKGNKNTKAE---CVRA 57

QY 95 YKPP 98

Db 58 CRPP 61

RESULT 2

TIBOBI

alpha-1-microglobulin / inter-alpha-trypsin inhibitor precursor [validated] - bovine

N:Alternate names: BT-14 (inhibitory fragment of ITI); bikunin; ITI

C:Species: Bos primigenius taurus (cattle)

C:Date: 25-Feb-1985 #sequence_revision 04-Feb-2000 #text_change 18-Aug-2000

C:Accession: S68149; A91717; A90685; S31219; A01209

R:Lindqvist, A.; Akerstrom, B.

Biochim. Biophys. Acta 1306, 98-106, 1996

A:Title: Bovine alpha(1)-microglobulin/bikunin. Isolation and characterization of liv

A:Reference number: S68149; MUID:96201710

A:Accession: S68149

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-352 <LIN>

A:Cross-references: EMBL:U35642; NID:g1016297; PIDN:AAB07599.1; PID:g1016298

R:Hochstrasser, K.; Wachter, E. Hoppe-Seyler's Z. Physiol. Chem. 364, 1679-1687, 1983 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-a A:Reference number: A91717; MUID:84133807 A:Accession: A91717 A:Molecule type: protein A:Residues: 227-267, 'L', 269-273, 'Q', 275-297, 'AF', 300-329, 'Q', 331-345, 'R', 347-348 <HOC> R:Hochstrasser, K.; Wachter, E.; Albrecht, G.J.; Relsinger, P. Biol. Chem. Hoppe-Seyler 366, 473-478, 1985 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-a A:Reference number: A90685; MUID:85225967 A:Accession: A90685 A:Molecule type: protein A:Residues: 347-349 <HOC2> R:Hochstrasser, K.; Albrecht, G.J.; Schonberger, O.L.; Wachter, E. Hoppe-Seyler's Z. Physiol. Chem. 364, 1689-1696, 1983 A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-a A:Reference number: A91718; MUID:84133808 A:Contents: annotation; reactive sites R:Castillo, G.M.; Templeton, D.M. FEBS Lett. 318, 292-296, 1993 A:Title: Subunit structure of bovine ESF (extracellular-matrix stabilizing factor(s)). A A:Reference number: S31219; MUID:93178646 A:Accession: S31219 A>Status: preliminary A:Molecule type: protein A:Residues: 206-214, 'X', 216, 'X', 218-220 <CAS> C:Superfamily: protein HC; animal Kunitz-type proteinase inhibitor homology; lipocalin h F:35-188/Domain: lipocalin homology <LIP> F:231-281/Domain: animal Kunitz-type proteinase inhibitor homology <BP1> F:287-337/Domain: animal Kunitz-type proteinase inhibitor homology <BP2> F:241/inhibitory site: leu (chymotrypsin, elastase) #status experimental F:250/Binding site: carbohydrate (asn) (covalent) #status experimental F:297/inhibitory site: Arg (trypsin) #status experimental	Db 1497 QNHKDACQLPKVPGCSGRKSHYVNTASHQCETFTYGGCLGNTRFATIECOARCPs 1556 Qy 94 KY 95 Db 1557 KF 1558 RESULT 4 S07451 proteinase inhibitor 5.II - snake-locks sea anemone C:Species: Anemonia sulcata (snake-locks sea anemone) C:Date: 31-Dec-1990 #sequence_revision 09-May-1997 #text_change 09-May-1997 C:Accession: S07451; B27222 R:Wunderer, G.; Machleidt, W.; Fritz, H. Meth. Enzymol. 80, 816-820, 1981 A:Title: The broad-specificity proteinase inhibitor 5 II from the sea anemone Anemoni A:Reference number: S07451 A:Accession: S07451 A:Molecule type: protein A:Residues: 1-59 <WUN> A:Note: 13-Arg, 16-Gly, 17-Gly, 25-Leu, 28-Arg, and 39-Arg were also found R:Krebs, H.C.; Habermehl, G.G. Naturwissenschaften 74, 395-396, 1987 A:Title: Isolierung und strukturaufklaerung eines haemolytisch aktiven peptids aus de A:Reference number: A94700 A:Accession: B27222 A:Molecule type: protein A:Residues: 1-38, 'R', 40, 'B', 42, 'BB', 45-48, 'ZZ', 51, 'Z', 53-62 <KRE> C:Superfamily: basic proteinase inhibitor; animal Kunitz-type proteinase inhibitor ho C:Keywords: serine proteinase inhibitor F:5-55/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>	Query Match 24.1%; Score 130; DB 2; Length 62; Best Local Similarity 45.8%; Pred. No. 1.4e-07; Matches 27; Conservative 5; Mismatches 23; Indels 4; Gaps 1; Qy 33 ICGDLKDPCKLDMNFGSCYEVHFRFYFNRTSKRCETFTVFGSCNGLNFKLKIEREVAC 91 Db 1 IINGD---CELPKVVGPCRARFPFYNNSSKRCCKFYGGCGNANNFHTLEECKVC 55 RESULT 5 HCHU alpha-1-microglobulin/inter-alpha-trypsin inhibitor precursor - human N:Alternate names: bikunin; complex-forming glycoprotein heterogeneous in charge (HC) rich protein N:Contains: alpha-1-microglobulin (protein HC); inter-alpha-trypsin inhibitor C:Species: Homo sapiens (man) C:Date: 15-Oct-1982 #sequence_revision 30-Jun-1987 #text_change 04-Feb-2000 C:Accession: S13433; S10778; A93642; A90074; A90225; A90686; PN0450; B39079; A61580; 3217 R:Veit, H.; Gebhard, W. Biol. Chem. Hoppe-Seyler 371, 1185-1196, 1990 A:Title: Structure of the human alpha(1)-microglobulin-bikunin gene. A:Reference number: S13433; MUID:91214554 A:Accession: S13433 A>Status: preliminary A:Molecule type: DNA A:Residues: 1-352 <VE1> A:Cross-references: EMBL:X54816; NID:g2475; PIDN:CAA3585.1; PID:g825614; EMBL:X5481 R:Diarr-Mehrpour, M.; Bourguignon, J.; Sesbouee, R.; Saller, J.P.; Levellard, T.; M Eur. J. Biochem. 191, 131-139, 1990 A:Title: Structural analysis of the human inter-alpha-trypsin inhibitor light-chain g A:Reference number: S10778; MUID:90336621 A:Accession: S10778 A>Status: preliminary A:Molecule type: DNA A:Residues: 1-202 <DIA> R:Kaumeyer, J.F.; Polazzi, J.O.; Kotick, M.P. Nucleic Acids Res. 14, 7839-7850, 1986 A:Title: The mRNA for a proteinase inhibitor related to the HI-30 domain of inter-alp
Qy 39 DPCKLDMNFGSCYEVHFRFYFNRTSKRCETFTVFGSCNGLNLF 81 Db 229 DSCOLDYSQPCGLGFRFYNGTSMACETFLYGGCMGNF 271 RESULT 3 T34394 hypothetical protein C37C3.6a - Caenorhabditis elegans C:Species: Caenorhabditis elegans C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999 C:Accession: T34394 R:Geisel, C.; Bradshaw, H. submitted to the EMBL Data Library, July 1996 A:Description: The sequence of C. elegans cosmid C37C3. A:Reference number: Z21518 A:Accession: T34394 A>Status: preliminary; translated from GB/EMBL/DBJ A:Molecule type: DNA A:Residues: 1-1558 <GE> A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a A:Experimental source: strain Bristol N2; clone C37C3 C:Genetics: A:Gene: CESP:C37C3.6a A:Map position: 5 A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1	Query Match 24.9%; Score 134; DB 1; Length 352; Best Local Similarity 55.8%; Pred. No. 2.7e-07; Matches 24; Conservative 5; Mismatches 14; Indels 0; Gaps 0; Qy 39 DPCKLDMNFGSCYEVHFRFYFNRTSKRCETFTVFGSCNGLNLF 81 Db 229 DSCOLDYSQPCGLGFRFYNGTSMACETFLYGGCMGNF 271 RESULT 3 T34394 hypothetical protein C37C3.6a - Caenorhabditis elegans C:Species: Caenorhabditis elegans C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999 C:Accession: T34394 R:Geisel, C.; Bradshaw, H. submitted to the EMBL Data Library, July 1996 A:Description: The sequence of C. elegans cosmid C37C3. A:Reference number: Z21518 A:Accession: T34394 A>Status: preliminary; translated from GB/EMBL/DBJ A:Molecule type: DNA A:Residues: 1-1558 <GE> A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a A:Experimental source: strain Bristol N2; clone C37C3 C:Genetics: A:Gene: CESP:C37C3.6a A:Map position: 5 A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1	Query Match 24.7%; Score 133; DB 2; Length 1558; Best Local Similarity 40.3%; Pred. No. 1.5e-06; Matches 25; Conservative 8; Mismatches 29; Indels 0; Gaps 0; Qy 34 CGDLKDPCKLDMNFGSCYEVHFRFYFNRTSKRCETFTVFGSCNGLNFKLKIEREVACVA 93

A:Reference number: A93642; MUID:87040757
 A:Accession: A93642
 A:Molecule type: mRNA
 A:Residues: 1-352 <KAD>
 A:Cross-references: GB:X04494; NID:g24478; PIDN:CAA28182.1; PID:g24479
 R:Lopez Otin, C.; Grubb, A.O.; Mendez, E.
 Arch. Biochem. Biophys. 228, 544-554, 1984
 A:Title: The complete amino acid sequence of human complex-forming glycoprotein heteroglycin
 A:Reference number: A90074; MUID:84126849
 A:Accession: A90074
 A:Molecule type: protein
 A:Residues: 20-56,58-202 <LOP>
 A:Experimental source: individual with tubular proteinuria
 A:Note: no evidence of sequence heterogeneity could be found, in spite of persistent heterogeneity
 R:Takagi, T.; Takagi, K.; Kawai, T.
 Biochem. Biophys. Res. Commun. 98, 997-1001, 1981
 A:Title: Complete amino acid sequence of human alpha-1-microglobulin.
 A:Reference number: A90225; MUID:81184038
 A:Accession: A90225
 A:Molecule type: protein
 A:Residues: 20-47;58-136,138-141,'T',143-144,146-198 <TAK>
 A:Experimental source: pooled urine of patients with tubular proteinuria
 R:Reisinger, P.; Hochstrasser, K.; Albrecht, G.J.; Lempert, K.; Salier, J.P.
 Biol. Chem. Hoppe-Seyler 366, 479-483, 1985
 A:Title: Human inter-alpha-trypsin inhibitor: localization of the kunitz-type domains in human plasma
 A:Reference number: A90686; MUID:85225968
 A:Accession: A90686
 A:Molecule type: protein
 A:Residues: 206-290, VI',293-342,'E',344-350 <REI>
 R:Atmani, F.; Lacour, B.; Strecker, G.; Parvy, P.; Druceke, T.; Daudon, M.
 Biochem. Biophys. Res. Commun. 191, 1158-1165, 1993
 A:Title: Molecular characteristics of uronic-acid-rich protein, a strong inhibitor of chondroitinase ABC
 A:Reference number: PN0450; MUID:93221481
 A:Accession: PN0450
 A:Molecule type: protein
 A:Residues: 206-214,'X' <ATM1>
 R:Englild, J.J.; Salvesen, G.; Hefta, S.A.; Thøgersen, I.B.; Rutherford, S.; Pizzo, S.V.
 J. Biol. Chem. 266, 747-751, 1991
 A:Title: Chondroitin 4-sulfate covalently cross-links the chains of the human blood protein chondroitin-6-sulfate
 A:Reference number: A39079; MUID:91093267
 A:Accession: B39079
 A:Molecule type: protein
 A:Residues: 206-225 <ENG1>
 R:Chirat, F.; Balduyck, M.; Mizon, C.; Laroui, S.; Sautiere, P.; Mizon, J.
 Int. J. Biochem. 23, 1201-1203, 1991
 A:Title: A chondroitin-sulfate chain is located on serine-10 of the urinary trypsin inhibitor
 A:Reference number: A61580; MUID:92175157
 A:Accession: A61580
 A:Molecule type: protein
 A:Residues: 214,'X',216-222,'X' <CHI>
 R:McKeehan, W.L.; Sakagami, Y.; Hoshi, H.; McKeehan, K.A.
 J. Biol. Chem. 261, 5378-5383, 1986
 A:Title: Two apparent human endothelial cell growth factors from human hepatoma cells and their cDNAs
 A:Reference number: A92583; MUID:86168278
 A:Accession: B25604
 A:Molecule type: protein
 A:Residues: 206-214,'X',216-230,'X',232-239,'X',241-248,'XX',251-252,'X',254 <MCK>
 R:Englild, J.J.; Thøgersen, I.B.; Pizzo, S.V.; Salvesen, G.
 J. Biol. Chem. 264, 15975-15981, 1989
 A:Title: Analysis of inter-alpha-trypsin inhibitor and a novel trypsin inhibitor, pre-alpha-1-trypsin inhibitor
 A:Reference number: A92736; MUID:89380192
 A:Accession: C34245
 A:Molecule type: protein
 A:Residues: 206-225 <ENG2>
 R:Traboni, C.; Cortese, R.
 Nucleic Acids Res. 14, 6340, 1986
 A:Title: Sequence of a full length cDNA coding for human protein HC (alpha-1-microglobulin)
 A:Reference number: A25303; MUID:86312901
 A:Accession: A25303
 A:Molecule type: mRNA
 A:Residues: 1-218,'HW' <TRA>
 A:Note: this mRNA sequence appears to contain errors after residue 218
 R:Calero, M.; Escribano, J.; Grubb, A.; Mendez, E.
 J. Biol. Chem. 269, 384-389, 1994
 A:Title: Location of a novel type of interpolypeptide chain linkage in the human protein HC
 A:Reference number: A53110; MUID:94103241
 A:Accession: A53110
 A:Molecule type: protein
 A:Residues: 45-57 <CAL1>
 R:Vetr, H.; Koegler, M.; Gebhard, W.
 FEBS Lett. 245, 137-140, 1989
 A:Title: The domain structure of the inhibitor subunit of human inter-alpha-trypsin I
 A:Reference number: S03552; MUID:89171290
 A:Accession: S03552
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 206-352 <VET2>
 R:Walki, N.; Balduyck, M.; Mases, P.; Capon, C.; Mizon, C.; Han, K.K.; Tartar, A.; Fou
 Biol. Chem. Hoppe-Seyler 373, 1009-1018, 1992
 A:Title: The heavy chains of human plasma inter-alpha-trypsin inhibitor: their isolat
 A:Reference number: S28928; MUID:93039735
 A:Accession: S28930
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 206-215 <MAL>
 R:Morelle, W.; Capon, C.; Balduyck, M.; Sautiere, P.; Kouach, M.; Michalski, C.; Four
 Eur. J. Biochem. 221, 881-888, 1994
 A:Title: Chondroitin sulphate covalently cross-links the three polypeptide chains of
 A:Reference number: S43466; MUID:94229087
 A:Accession: S43466
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 206-221 <MOR>
 R:Wisniewski, H.G.; Burgess, W.H.; Oppenheim, J.D.; Vilcek, J.
 Biochemistry 33, 7423-7429, 1994
 A:Title: TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable co
 A:Reference number: A53642; MUID:94271799
 A:Accession: A53642
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 206-217 <WIS>
 R:Calero, M.; Mendez, E.; Garcia, E.
 Biochim. Biophys. Acta 1249, 91-99, 1995
 A:Title: Expression of the human complex-forming glycoprotein HC (alpha-1-microglobul
 A:Reference number: S55688; MUID:95284116
 A:Accession: S55688
 A:Molecule type: protein
 A:Residues: 20-24 <CAL2>
 R:Bourguignon, J.; Diarra-Mehrpour, M.; Sesboue, R.; Frain, M.; Sala-Trepas, J.M.; Ma
 Biochem. Biophys. Res. Commun. 131, 1146-1153, 1985
 A:Title: Human inter-alpha-trypsin-inhibitor: characterization and partial nucleotide
 A:Reference number: I52208; MUID:86025577
 A:Accession: I52208
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 302-352 <BOU>
 A:Cross-references: GB:M11562; NID:g186587; PIDN:AAA59194.1; PID:g307077
 R:Wojsik, E.G.C.; van den Berg, M.; van der Linden, I.K.; Poort, S.R.; Cupers, R.; Be
 Biochem. J. 311, 753-759, 1995
 A:Title: Factor IX zutphen: a Cys(18) -> Arg mutation results in formation of a heter
 A:Reference number: S59509; MUID:96067589
 A:Accession: S59509
 A:Molecule type: protein
 A:Residues: 27-35,'Y',37 <WOJ>
 R:Atmani, F.; Mizon, J.; Khan, S.R.
 Eur. J. Biochem. 236, 984-990, 1996
 A:Title: Identification of uronic-acid-rich protein as urinary bikunin, the light cha
 A:Reference number: S66434; MUID:96270753
 A:Accession: S66434
 A:Molecule type: protein
 A:Residues: 206-214,'X',216-230 <ATM2>
 R:Akerstroem, B.; Bratt, T.; Englild, J.J.
 FEBS Lett. 362, 50-54, 1995
 A:Title: Formation of the alpha(1)-microglobulin chromophore in mammalian and insect
 A:Reference number: S68728; MUID:95212582
 A:Accession: S68728

A:Molecule type: protein	
A:Residues: 89-100 <AKE>	
R:Jessen, T.E.; Faarvang, K.L.; Ploug, M.	
FEBS Lett. 230, 195-200, 1988	
A:Title: Carbohydrate as covalent crosslink in human inter-alpha-trypsin inhibitor: a no	
A:Reference number: S02431; MUID:88167187	
A:Accession: S02431	
A:Molecule type: protein	
A:Residues: 206-214, 'X', 216-217 <JES>	
R:Lopez, C.; Grubb, A.; Mendez, E.	
FEBS Lett. 144, 349-353, 1982	
A:Title: Human protein HC displays variability in its carboxyl-terminal amino acid sequ	
A:Reference number: A91304	
A:Contents: annotation; variant of alpha-1-microglobulin	
A>Note: pooled urine samples contained two forms of this protein, both lacking 57-Lys an	
R:Hochstrasser, K.; Schonberger, O.L.; Rossmann, I.; Wachter, E.	
Hoppe-Seyler's Z. Physiol. Chem. 362, 1357-1362, 1981	
A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-a	
by affinity chromatography: MUID:82074265	
A:Reference number: A91698	
A:Contents: annotation; carbohydrate binding sites	
R:Morii, M.; Travis, J.	
Biol. Chem. Hoppe-Seyler 366, 19-21, 1985	
A:Title: The reactive site of human inter-alpha-trypsin inhibitor is in the amino-termi	
A:Reference number: A90682; MUID:85225940	
A:Contents: annotation; inhibitory site	
A>Note: in vitro, the first twelve residues of the amino end of the inhibitor appear to	
wn	
C:Comment: Alpha-1-microglobulin and inter-alpha-trypsin inhibitor are proteolytically p	
C:Comment: Alpha-1-microglobulin occurs in many physiological fluids including plasma, u	
. It contains at least one brown-yellow chromophore.	
Query Match	24.0%; Score 129.5; DB 1; Length 352;
Best Local Similarity	41.4%; Pred. No. 8.6e-07;
Matches	29; Conservative 9; Mismatches 25; Indels 7; Gaps 2;
QY	24 GG--VNKIAKIGCDLKDPCKLDMNFGSCYEVHFRFYNRTSKRCETFFVFGCGNGLNLF 81
Db	217 GGLVTEVTKK-----EDSCQLGYSAGPCMGWTSRYFYNGTSMACETFYQGCMGNGNF 271
QY	82 KLKIEREVAC 91
Db	272 VTEKECLQTC 281
RESULT 6	
T34395	
hypothetical protein C37C3.6b - Caenorhabditis elegans	
C:Species: Caenorhabditis elegans	
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999	
C:Accession: T34395	
R:Geisel, C.; Bradshaw, H.	
submitted to the EMBL Data Library, July 1996	
A:Description: The sequence of C. elegans cosmid C37C3.	
A:Reference number: Z21518	
A:Accession: T34395	
A>Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-2167 <GEI>	
A:Cross-references: EMBL:U64857; PIDN:AC25868.1; GSPDB:GN00023; CESP:C37C3.6b	
A:Experimental source: strain Bristol N2; clone C37C3	
C:Genetics:	
A:Gene: CESP:C37C3.6b	
A:Map position: 5	
A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 1556/	
Query Match	23.7%; Score 128; DB 2; Length 2167;
Best Local Similarity	41.8%; Pred. No. 7.6e-06;
Matches	23; Conservative 9; Mismatches 23; Indels 0; Gaps 0;
QY	37 LKDPCKLDMNFGSCYEVHFRFYNRTSKRCETFFVFGCGNGLNFKLKIEREVAC 91
Db	272 VTEKECLQTC 281

Db	1267 MEDICRSRQDAGPCETYSQWFFYNAFSQECETFTYGGCGNGLNFRSKDECEQRC 1321
RESULT 7	
TIVIVC	
venom basic proteinase inhibitor III - sand viper	
N:Alternate names: venom chymotrypsin inhibitor	
C:Species: Vipera ammodytes (sand viper)	
C:Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 16-Aug-1996	
C:Accession: A01223	
R:Ritonja, A.; Meloun, B.; Gubensek, F.	
Biochim. Biophys. Acta 746, 138-145, 1983	
A:Title: The primary structure of Vipera ammodytes venom chymotrypsin inhibitor.	
A:Reference number: A01223	
A:Accession: A01223	
A:Molecule type: protein	
A:Residues: 1-65 <RII>	
C:Superfamily: basic proteinase inhibitor; animal Kunitz-type proteinase inhibitor ho	
C:Keywords: serine proteinase inhibitor; venom	
F:7-57/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>	
F:7-57.16-40.32-53/Disulfide bonds: #status predicted	
F:17/Inhibitory site: Leu (chymotrypsin) #status predicted	
Query Match	23.6%; Score 127; DB 1; Length 65;
Best Local Similarity	42.4%; Pred. No. 3e-07;
Matches	25; Conservative 6; Mismatches 28; Indels 0; Gaps 0;
QY	41 CKLDMNFGSCYEVHFRFYNRTSKRCETFFVFGCGNGLNFKLKIEREVACVAKYKPPR 99
Db	7 CYLPADPGRCGLAYMPRFYFNPSANKCEFIYGGCGNANNFTWDECHITCVASGIQPR 65
RESULT 8	
A29652	
inter-alpha-trypsin inhibitor (BPI type) - sheep (fragment)	
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)	
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 16-Jul-1999	
C:Accession: A29652	
R:Rasp, G.; Hochstrasser, K.; Wachter, E.; Reisinger, P.W.M.	
Biol. Chem. Hoppe-Seyler 368, 727-731, 1987	
A:Title: The amino-acid sequence of the trypsin-released inhibitor from sheep inter-a	
sin inhibitor, XI.)	
A:Reference number: A29652; MUID:87299012	
A:Accession: A29652	
A:Molecule type: protein	
A:Residues: 1-123 <RAS>	
C:Superfamily: protein HC; animal Kunitz-type proteinase inhibitor homology; lipocali	
C:Keywords: serine proteinase inhibitor	
F:5-55/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>	
F:61-111/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>	
Query Match	23.6%; Score 127; DB 2; Length 123;
Best Local Similarity	46.3%; Pred. No. 5.7e-07;
Matches	25; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
QY	38 KDPCKLDMNFGSCYEVHFRFYNRTSKRCETFFVFGCGNGLNFKLKIEREVAC 91
Db	2 EDSCQLGYSQGFCLGMFRFYNGTSMACETFYGGCGNGLNFPSEKELQTC 55
RESULT 9	
TIHOBI	
alpha-1-microglobulin/inter-alpha-trypsin inhibitor - horse (fragment)	
N:Alternate names: EI-14 (inhibitory fragment of ITI); ITI; trypsin inhibitor, E-UTI	
C:Species: Equus caballus (domestic horse)	
C:Date: 30-Jun-1987 #sequence_revision 04-Feb-2000 #text_change 05-May-2000	
C:Accession: A01210; A45653	
R:Hochstrasser, K.; Wachter, E.; Albrecht, G.J.; Reisinger, P.	
Biol. Chem. Hoppe-Seyler 366, 473-478, 1985	
A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inte	
A:Reference number: A90685; MUID:85225967	

Query Match	23.4%	Score 126;	DB 1;	Length 67;
Best Local Similarity	41.0%	Pred. No. 4e-07;		
Matches 25;	Conservative	6;	Mismatches 26;	Indels
				Gaps 1;

Db 180 ANNFQ 184

|||||

RESULT 13

A54951

N:Alternate names: placental inhibitor-2 precursor - human

C:Species: Homo sapiens (man)

C>Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 16-Jul-1999

C:Accession: A54951; I55185; A34029; C34029; B34029

R:Sprecher, C.A.; Kistiel, W.; Mathewes, S.; Foster, D.C.

Proc. Natl. Acad. Sci. U.S.A. 91, 3353-3357, 1994

A:Title: Molecular cloning, expression, and partial characterization of a second human

A:Reference number: A54951; MUID:942111862

A:Accession: A54951

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-235 <RES>

A:Cross-references: GB:L27624; NID:9441149; PIDN:AAA20094.1; PID:9441150

A:Experimental source: Placenta

R:Miyagi, Y.; Koshikawa, N.; Yasumitsu, H.; Miyagi, E.; Hirahara, F.; Aoki, I.; Misugi, J. Biochem. 116, 939-942, 1994

A:Title: cDNA cloning and mRNA expression of a serine proteinase inhibitor secreted by

A:Reference number: I55185; MUID:95204397

A:Accession: I55185

A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL

A:Molecule type: mRNA

A:Residues: 1-235 <RES>

A:Cross-references: GB:D29992; NID:9484050; PIDN:BAA06272.1; PID:9484051

A:Note: parts of this sequence, including the amino end of the mature protein, were dete

R:Buetzow, R.; Huhtala, M.L.; Bohn, H.; Virtanen, I.; Seppaelae, M.

Biochem. Biophys. Res. Commun. 150, 483-490, 1988

A:Title: Purification and characterization of placental protein 5.

A:Reference number: A34029; MUID:88106628

A:Accession: A34029

A:Molecule type: protein

A:Residues: 'A', 24-33, 'X', 35 <BU2>

A:Accession: C34029

A:Molecule type: protein

A:Residues: 47-50, 'X', 52-53 <BU2>

A:Accession: B34029

A:Molecule type: protein

A:Residues: 133, 'X', 135-137, 'X', 139-140, 'X', 142-144, 'X', 146 <BU3>

C:Genetics:

A:Gene: GDB:TFP12

A:Cross-references: GDB:354485

C:Superfamily: tissue factor pathway inhibitor; animal Kunitz-type proteinase inhibitor

C:Keywords: anticoagulant; blood coagulation; duplication; glycoprotein; heparin binding

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-235/Product: tissue factor pathway inhibitor-2 #status predicted <MAT>

F:36-86/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

F:96-149/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>

F:158-208/Domain: animal Kunitz-type proteinase inhibitor homology <BP3>

F:36-86, 45-69, 61-82, 96-149, 106-130, 122-145, 158-208, 167-191, 183-204/Disulfide bonds: #sta

Query Match 23.0%; Score 124; DB 2; Length 235;

Best Local Similarity 31.4%; Pred. No. 2.3e-06;

Matches 27; Conservative 18; Mismatches 37; Indels 4; Gaps 1;

QY 18 LNTLLG----GGVNKIAETGCKDLKDPCKLDMNFGSCYEVHFRFYFNRTSKRCETFFVFSG 73

Db 9 LSILLFLTEALGDAQEPFGNNAETCLLPIDYGPCRALLRLRYDYRTQSCRQFLYGG 68

QY 74 CNGNLFNFKLIEREVACVAKYKPPR 99

Db 69 CEGNANNFYTWACDDACWRIKVPK 94

RESULT 14

S35708

alpha-1-microglobulin / bikunin precursor - mouse

N:Alternate names: alpha 1-microglobulin / inter-alpha-trypsin inhibitor light chain

C:Species: Mus musculus (house mouse)

C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Jun-2000

C:Accession: S35708; JX0355

R:Chan, P.; Saller, J.P.

Biochim. Biophys. Acta 1174, 195-200, 1993

A:Title: Mouse alpha-1-microglobulin/bikunin precursor: cDNA analysis, gene evolution

A:Reference number: S35708; MUID:93363639

A:Accession: S35708

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-349 <CHA>

A:Cross-references: EMBL:X68680; NID:9311702; PIDN:CAA48640.1; PID:9311703

R:Itoh, H.; Ide, H.; Kataoka, H.; Tomita, M.; Yoshihara, H.; Nawa, Y.

J. Biochem. 116, 767-772, 1994

A:Title: cDNA sequencing of mouse alpha1-microglobulin/inter-alpha-trypsin inhibitor

A:Reference number: JX0355; MUID:95189774

A:Accession: JX0355

A:Molecule type: mRNA

A:Residues: 1-64, 'S', 66-349 <ITO>

A:Cross-references: DDBJ:D28812; NID:9556530; PIDN:BAA05973.1; PID:9556531

A:Experimental source: liver

C:Superfamily: protein HC; animal Kunitz-type proteinase inhibitor homology; lipocal

C:Keywords: glycoprotein; inflammation; serine proteinase inhibitor

F:34-187/Domain: lipocalin homology <LIP>

F:230-280/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

F:286-336/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>

F:114,233/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.0%; Score 124; DB 2; Length 349;

Best Local Similarity 49.0%; Pred. No. 3.4e-06;

Matches 25; Conservative 6; Mismatches 18; Indels 2; Gaps 1;

QY 33 ICGDLK--DPCKLDMNFGSCYEVHFRFYFNRTSKRCETFFVFSGCGNLLNF 81

Db 220 ITGTLKEDSCOLYSEGCPLGMQERYIYNGASMACETFFYGGCLGNGNFF 270

RESULT 15

T26063

hypothetical protein W01F3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T26063

R:Cummings, P.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z20145

A:Accession: T26063

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2225 <WIL>

A:Cross-references: EMBL:T292815; PIDN:CAB07294.1; GSPDB:GN00023; CESP:W01F3.3

A:Experimental source: clone W01F3

C:Genetics:

A:Gene: CESP:W01F3.3

A:Map position: 5

A:Introns: 33/1; 56/1; 100/1; 142/3; 271/3; 451/1; 525/3; 774/1; 1093/1; 1178/1; 1221

Query Match 22.9%; Score 123.5; DB 2; Length 2225;

Best Local Similarity 35.4%; Pred. No. 2.4e-05;

Matches 28; Conservative 12; Mismatches 32; Indels 7; Gaps 2;

QY 19 NTLILGGVNKIAETGCKDLKDPCKLDMNFGSCYEVHFRFYFNRTSKRCETFFVFSGCGNGL 78

Db 759 NTVLLGGI-----EDTTTDSVNRCLHPRDSGNCRGQVRFVFEDEKKKNDVFTYTCGCGN 814

QY 79 NNFKLKIEREVACVAKYKPP 97

Db 815 NNFASKECMCAIC---HKP 830

Search completed: March 24, 2001, 13:23:14
Job time: 104 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:23:51 ; Search time 34.59 Seconds
(without alignments)
51.395 Million cell updates/sec

Title: US-09-215-435-177

Perfect score: 539

Sequence: 1 MSAKLGFLRRFFIFCSLNT.....NFKLKIERVACVAKYKPPR 99

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2.6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2.6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2.6/ptodata/2/1aa/6_COMB.pep:*
4: /cgn2.6/ptodata/2/1aa/PTUS_COMB.pep:*
5: /cgn2.6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154	28.6	66	1 US-08-358-160-113	Sequence 113, App
2	154	28.6	79	5 5466783-7	Patent No. 5466783
3	142	26.3	60	1 US-08-358-160-153	Sequence 153, App
4	142	26.3	122	2 US-08-422-333-12	Sequence 12, App
5	142	26.3	122	5 5187153-20	Patent No. 5187153
6	142	26.3	122	5 5220013-23	Patent No. 5220013
7	135	25.0	58	1 US-08-676-125A-46	Sequence 46, App
8	135	25.0	58	2 US-09-136-012A-46	Sequence 46, App
9	132	24.5	58	3 US-08-676-124-65	Sequence 65, App
10	132	24.5	58	3 US-09-414-878-65	Sequence 65, App
11	132	24.5	58	3 US-09-240-136-65	Sequence 65, App
12	131	24.3	58	3 US-08-676-124-77	Sequence 77, App
13	131	24.3	58	3 US-09-414-878-77	Sequence 77, App
14	131	24.3	58	3 US-09-240-136-77	Sequence 77, App
15	131	24.3	484	1 US-08-358-160-7	Sequence 7, App
16	130	24.1	62	1 US-08-358-160-97	Sequence 97, App
17	129.5	24.0	143	2 US-08-422-333-10	Sequence 10, App
18	129.5	24.0	143	5 5223482-20	Patent No. 5223482
19	129.5	24.0	144	5 5187153-18	Patent No. 5187153
20	129.5	24.0	147	1 US-08-358-160-72	Sequence 72, App
21	129	23.9	56	1 US-08-358-160-87	Sequence 87, App
22	129	23.9	67	1 US-08-358-160-123	Sequence 123, App
23	128.5	23.8	122	5 5466783-23	Patent No. 5466783
24	128	23.7	58	1 US-08-358-160-17	Sequence 17, App
25	127	23.6	58	1 US-08-384-489-10	Sequence 10, App
26	127	23.6	58	1 US-08-358-160-86	Sequence 86, App
27	127	23.6	58	1 US-08-463-155A-38	Sequence 38, App
28	127	23.6	58	1 US-08-463-432B-38	Sequence 38, App

ALIGNMENTS

RESULT 1

US-08-358-160-113

; Sequence 113, Application US/08358160

; Patent No. 5663143

; GENERAL INFORMATION:

; APPLICANT: LEY, Arthur C.

; APPLICANT: LADNER, Robert C.

; APPLICANT: GUTERMAN, Sonia K.

; APPLICANT: ROBERTS, Bruce L.

; APPLICANT: MARKLAND, William

; APPLICANT: KENT, Rachel B

; TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ

; TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE

; NUMBER OF SEQUENCES: 234

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W. Suite 300

; CITY: Washington

; STATE: District of Columbia

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/358,160

; FILING DATE: 16-DEC-1994

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/133,031

; FILING DATE: 13-OCT-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/009,319

; FILING DATE: 26-JAN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/664,989

; FILING DATE: 01-MAR-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/487,063

; FILING DATE: 02-MAR-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/240,160

; FILING DATE: 02-SEP-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Cooper, Iver P.

; REGISTRATION NUMBER: 28,005

; REFERENCE/DOCKET NUMBER: LEY-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

Sequence 45, Appl
Sequence 38, Appl
Sequence 38, Appl
Sequence 38, Appl
Sequence 38, Appl
Sequence 45, Appl
Sequence 76, Appl
Sequence 76, Appl
Sequence 128, App
Sequence 132, App
Sequence 92, Appl
Sequence 18, Appl
Patent No. 5466783
Patent No. 5466783
Sequence 10, Appl
Sequence 10, Appl

58 1 US-08-676-125A-45
58 1 US-08-206-310A-38
58 2 US-08-398-010A-38
58 2 US-08-398-628A-38
58 2 US-08-399-115A-38
58 2 US-09-136-012A-45
58 3 US-08-676-124-76
58 3 US-09-414-878-76
58 3 US-09-240-136-76
60 1 US-08-358-160-128
60 1 US-08-358-160-132
65 1 US-08-358-160-92
111 1 US-08-321-658B-18
123 5 5466783-21
127 5 5466783-24
56 1 US-07-700-526-10
56 4 PCT-US92-03132-10

29 127 23.6
30 127 23.6
31 127 23.6
32 127 23.6
33 127 23.6
34 127 23.6
35 127 23.6
36 127 23.6
37 127 23.6
38 127 23.6
39 127 23.6
40 127 23.6
41 127 23.6
42 127 23.6
43 127 23.6
44 126 23.4
45 126 23.4

```

; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-358-160-113
; NUMBER OF SEQUENCES: 234
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W. Suite 300
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 16-DEC-1994
; PRIORITY APPLICATION NUMBER: US 08/358,160
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; FILING DATE: 13-OCT-1993
; APPLICATION NUMBER: US 08/133,031
; FILING DATE: 01-MAR-1991
; APPLICATION NUMBER: US 07/664,989
; PRIOR APPLICATION DATA:
; FILING DATE: 02-MAR-1990
; APPLICATION NUMBER: US 07/240,160
; FILING DATE: 02-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Cooper, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: LEY-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-358-160-153
; Query Match 26.3%; Score 142; DB 1; Length 60;
; Best Local Similarity 58.1%; Pred. No. 3.4e-10;
; Matches 25; Conservative 5; Mismatches 13; Indels 0; Gaps 0;
QY 39 DPCKLDNMGSCYEYHFVRYFNRTSKRCETVFSGCGNLNNF 81
DB 3 DSCQLDYSQGPGCLGLFKRYFYNGTSMACETFLYGCMGNLNF 45
;
RESULT 4
US-08-422-333-12
; Sequence 12, Application US/08422333
; Patent No. 5912410
; GENERAL INFORMATION:
; APPLICANT: CORDELL, Barbara L.
; TITLE OF INVENTION: TRANSGENIC NON-HUMAN MAMMAL DISPLAYING
; THE AMYLOID-FORMING PATHOLOGY OF ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scios, Inc.
; STREET: 2450 Bayshore Parkway
; CITY: Mountain View
; STATE: CA

```



```

; MOLECULE TYPE: protein
US-09-136-012A-46

Query Match          25.0%; Score 135; DB 2; Length 58;
Best Local Similarity 48.1%; Pred. NO. 2.2e-09;
Matches 26; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY    38 KDPCKLDMNFGSCYEYHFHYFNRTSKRCETEFVSGCGNGNLNNFKLKIEREVAC 91
      :|:|: |:|: | | | | | | | | | | | | | | | | | | | | | |
DB     2 EDCQLGYDAGPCRGAPRVFYNGTSMACETFSYGCGGCGNNNFVTEKELQTC 55

RESULT   9
US-08-676-124-65
; Sequence 65, Application US/08676124
; Patent No. 6010880
; GENERAL INFORMATION:
; APPLICANT: MARKLAND, William
; APPLICANT: LADNER, Robert Charles
; TITLE OF INVENTION: INHIBITORS OF HUMAN PLASMIN DERIVED
; TITLE OF INVENTION: FROM FROM THE KUNITZ DOMAINS
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,124
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00298
; FILING DATE: 11-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,658
; FILING DATE: 11-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/208,265
; FILING DATE: 10-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: MARKLAND-3B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-676-124-65

Query Match          24.5%; Score 132; DB 3; Length 58;
Best Local Similarity 45.3%; Pred. NO. 5.1e-09;
Matches 24; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY    39 DPCKLDMNFGSCYEYHFHYFNRTSKRCETEFVSGCGNGLNANKLKIEREVAC 91
      | |:|: |:|: | | | | | | | | | | | | | | | | | |
DB     3 DFCLEEDTGPGRGRDFRYFNNOQTQCECTFIYGGCEGNMNNFTLECKNIC 55

```

```

; Sequence 65, Application US/09240136
; Patent No. 6103499
; GENERAL INFORMATION:
; APPLICANT: DYAX CORP
; APPLICANT: MARKLAND, William
; APPLICANT: LADNER, Robert C
; TITLE OF INVENTION: Inhibitors of Human Plamin Derived
; TITLE OF INVENTION: From The Kunitz Domains
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yankwich & Associates
; STREET: 130 Bishop Allen Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02139
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5-inch diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Microsoft Windows 98
; SOFTWARE: Microsoft Word 97 SR-1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/240,136
; FILING DATE: (concurrently herewith)
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/676,124
; FILING DATE: 07-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00298
; FILING DATE: 11-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/208,265
; FILING DATE: 10-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/179,685
; FILING DATE: 11-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: YANKWICH, Leon R
; REGISTRATION NUMBER: 30,237
; NAME: ZWICKER, Kenneth P
; REGISTRATION NUMBER: 43,310
; REFERENCE/DOCKET NUMBER: DYX-007.2P US-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-491-4343
; TELEFAX: 617-491-8801
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-240-136-65

Query Match 24.5%; Score 132; DB 3; Length 58;
Best Local Similarity 45.3%; Pred. No. 5,1e-09;
Matches 24; Conservative 9; Mismatches 20; Indels 0

QY 39 DPCKIDMFGSCYCYEHFRFYNYNTSKRCETFFVSGCGNGLNNFKLKIEREVAC 91
| | : | | | | | : | | | | : | | | | : | | | | : | |
DB 3 DFCFLEEDTGPCRGFRDIFYNNQTQCEFFIYGCBCGNMNFETLECKNIC 55

RESULT 12
US-08-676-124-77
; Sequence 77, Application US/08676124
; Patent No. 6010880
; GENERAL INFORMATION:
; APPLICANT: MARKLAND, William
; APPLICANT: LADNER, Robert Charles
; TITLE OF INVENTION: INHIBITORS OF HUMAN PLASMIN DERIVED

```

```

RESULT 13
US-09-414-878-77
; Sequence 77, Application US/09414878
; Patent No. 6071723
; GENERAL INFORMATION:
; APPLICANT: DYAX CORP
; APPLICANT: MARKLAND, William
; APPLICANT: LADNER, Robert C
; TITLE OF INVENTION: Inhibitors of Human Plamin Derived
; TITLE OF INVENTION: From The Kunitz Domains
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yankwich & Associates
; STREET: 130 Bishop Allen Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02139
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5-inch diskette

```

```

RESULT 14
US-09-240-77
Sequence 77, Application US/09240136
Patent No. 6103499
GENERAL INVENTION:
APPLICANT: DYAX CORP
APPLICANT: MARKLAND, William
APPLICANT: LADNER, Robert C
TITLE OF INVENTION: Inhibitors o
TITLE OF INVENTION: From The Kun
NUMBER OF SEQUENCES: 139
CORRESPONDENCE ADDRESS:
ADDRESS: Yankwich & Associat
STREET: 130 Bishop Allen Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02139
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5-inch diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Microsoft Win
SOFTWARE: Microsoft word 97 SR

```



```

CURRENT APPLICATION DATA:
- APPLICATION NUMBER: US/09/240,136
- FILING DATE: (concurrently herewith)
CLASSIFICATION:
PRIOR APPLICATION DATA:
- APPLICATION NUMBER: 08/676,124
- FILING DATE: 07-JAN-1997
PRIOR APPLICATION DATA:
- APPLICATION NUMBER: PCT/US95/00298
- FILING DATE: 11-JAN-1995
PRIOR APPLICATION DATA:
- APPLICATION NUMBER: 08/208,265
- FILING DATE: 10-MAR-1994
PRIOR APPLICATION DATA:
- APPLICATION NUMBER: 08/179,685
- FILING DATE: 11-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: YANKWICH, Leon R
REGISTRATION NUMBER: 30,237
NAME: ZWICKER, Kenneth P
REGISTRATION NUMBER: 43,310
REFERENCE/DOCKET NUMBER: DYX-007.2P US-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-491-4343
TELEFAX: 617-491-8801
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-240-136-77

```

```

Query Match      24.3%; Score 131; DB 3; Length 58;
Best Local Similarity 48.1%; Pred. No. 6.7e-09;
Matches 26; Conservative 4; Mismatches 24; Indels 0;

QY 38 KDPCKLDMNFGSCVEYHFRFYFNFTSKRCETFFVSGCGNLLNNFKLIEREVAC 91
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
db 2  EDSCOLCYEAGPCRGKTSFYFNFTSKRCETFFVSGCGNLLNNFKLIEREVAC 55
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 15
US-08-358-160-7
; Sequence 7, Application US/08358160
; Patent No. 563143
; GENERAL INFORMATION:
; APPLICANT: LEV, Arthur C.
; APPLICANT: LADNER, Robert C.
; APPLICANT: GUTERMAN, Sonia K.
; APPLICANT: ROBERTS, Bruce L.
; APPLICANT: MARKLAND, William
; APPLICANT: KENT, Rachel B.
; TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
; TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
; NUMBER OF SEQUENCES: 234
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W. Suite 300
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,160
; FILING DATE: 16-DEC-1994

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:21:34 ; Search time 25.25 Seconds
(without alignments)
448.920 Million cell updates/sec

Title: US-09-215-435-167
Perfect score: 1878
Sequence: 1 MYPFIYLAHFHLCGSWSST.....EGPFCHLGHRLWLQHSR 351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1673	89.1	501	BACE_HUMAN	P56817 homo sapien
2	1661	88.4	501	BACE_MOUSE	P56818 mus musculus
3	1655	88.1	501	BACE_RAT	P56819 rattus norv
4	925	49.3	518	BAE2_HUMAN	Q9Y520 homo sapien
5	259	13.8	324	PEP1_GADMO	P56272 gadus morhu
6	253.5	13.5	367	PEP1_CHICK	P00793 gallus gall
7	251	13.4	388	PEP2_MACFU	P27678 macaca fusc
8	239	12.7	387	PEP1_RABIT	P28712 oryctolagus
9	237.5	12.6	387	PEP3_RABIT	P27822 oryctolagus
10	234.5	12.5	388	PEP2_MACFU	P27677 macaca fusc
11	234	12.5	386	PEP1_PIG	P00791 sus scrofa
12	233	12.4	398	CATE_RAT	P16228 rattus norv
13	231.5	12.3	388	PEP1_HUMAN	P00790 homo sapien
14	231.5	12.3	388	PEP1_MACMU	P11489 macaca mula
15	230.5	12.3	383	PEPE_CHICK	P16476 gallus gall
16	230	12.2	387	PEP2_RABIT	P27821 oryctolagus
17	230	12.2	396	CATE_HUMAN	P14091 homo sapien
18	229.5	12.2	388	PEP1_MACFU	P03954 macaca fusc
19	229.5	12.2	391	CATE_CAVPO	P25796 cavia porce
20	229	12.2	387	PEP1_RABIT	P28713 oryctolagus
21	227	12.1	381	CHYM_SHEEP	P18276 ovis aries
22	224	11.9	412	CATD_HUMAN	P07339 homo sapien
23	223.5	11.9	388	PAG_HORSE	Q28389 equus cabal
24	220.5	11.7	388	PEPE_RABIT	P27823 oryctolagus
25	220	11.7	396	CATE_RABIT	P43159 oryctolagus
26	220	11.7	397	CATE_MOUSE	P70269 mus musculus
27	219	11.7	381	CHYM_BOVIN	P00794 bos taurus
28	218	11.6	377	PEPC_MACFU	P03955 macaca fusc
29	216.5	11.5	407	CATD_RAT	P24268 rattus norv
30	212	11.3	410	CATD_MOUSE	P18242 mus musculus
31	210	11.2	398	CATD_CHICK	Q05744 gallus gall
32	209	11.1	388	PEPC_HUMAN	P20142 homo sapien
33	202	10.8	394	PEPC_CAVPO	Q64411 cavia porce

34	200	10.6	419	1	CARV_CANAL	P10977 candida alb
35	197.5	10.5	345	1	CATD_PIG	P00795 sus scrofa
36	196	10.4	376	1	PAG2_BOVIN	Q28057 bos taurus
37	192.5	10.3	389	1	PAG1_PIG	Q29078 sus scrofa
38	184.5	9.8	402	1	REN1_RAT	P08424 rattus norv
39	183.5	9.8	392	1	PEPC_RAT	Q29073 rattus norv
40	183.5	9.8	420	1	PAG2_PIG	Q29079 sus scrofa
41	182	9.7	380	1	PAG1_BOVIN	Q29432 bos taurus
42	178.5	9.5	387	1	ASPP_AEDAE	Q03168 aedes aegyp
43	178	9.5	396	1	CARP_NEUCR	O01294 neuropept
44	175.5	9.3	394	1	PEPA_ASFAW	P17946 aspergillus
45	175	9.3	496	1	ASPR_ORYSA	P42211 oryza sativ

ALIGNMENTS

RESULT 1
BACE_HUMAN
ID BACE_HUMAN STANDARD; PRT; 501 AA.
AC P56817; Q9UJT5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE BETA-SECRETASE PRECURSOR (EC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)
DE (BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL
DE PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC PROTEASE 2)
DE (MEMAPIN-2);
GN BACE OR BACE1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=20002972; PubMed=10531052;
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,
RA Treanor J., Rogers G., Citron M.;
RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
RT the transmembrane aspartic protease BACE.";
RN Science 286:735-741(1999).
[2]
RP SEQUENCE FROM N.A., SEQUENCE OF 46-68, AND CHARACTERIZATION
RC TISSUE-BRAIN;
RX MEDLINE=20057171; PubMed=10591214;
RA Sinha S., Anderson J.P., Barbour R., Basi G.S., Caccavello R.,
RA Davis D., Doan M., Dovey H.F., Frigon J., Hong J., Jacobson-Croak K.,
RA Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H.,
RA Tatsuno G., Tung J., Schenk D., Seubert P., Somersari S.M., Wang S.,
Walker D., Zhao J., McConlogue L., Varghese J.;
RT "Purification and cloning of amyloid precursor protein beta-secretase
RT from human brain.";
RN Nature 402:537-540(1999).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057170; PubMed=10591213;
RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
RA Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;
RT "Membrane-anchored aspartyl protease with Alzheimer's disease beta-
RT secretase activity.";
RN Nature 402:533-537(1999).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20030166; PubMed=10561122;
RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,
RA Chapman C., Glozier I.S., Murphy K.E., Southan C.D., Ryan D.M.,
RA Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;
RT "Identification of a novel aspartic proteinase (Asp 2) as beta-
RT secretase.";

Mol. Cell. Neurosci. 14:419-427(1999).

[51]
SEQUENCE OF 14-501 FROM N.A., AND CHARACTERIZATION.
MEDLINE=20144060; PubMed=10677483;
Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;
"Human aspartic protease memapsin 2 cleaves the beta-secretase site of
beta-amyloid precursor protein.";
Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).
-!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP.
LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE.
-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
-!- TISSUE SPECIFICITY: BRAIN.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
EUKARYOTIC ASPARTYL PROTEASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AF190725; AAF04142.1; -
CC EMBL; AF201468; AAF18982.1; -
CC EMBL; AF200343; AAF17079.1; -
CC EMBL; AF204943; AAF26367.1; -
CC EMBL; AF200193; AAF13715.1; -
CC MIN; 604252; -
CC INTERPRO; IPR001461; -
CC PFAM; PF00026; asp; 3.
CC PRINTS; PR00792; PEPsin.
CC PROSITE; PS00141; ASP_PROTEASE; 1.
KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;
Signal.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 45
FT CHAIN 46 501
FT DOMAIN 22 457
FT TRANSMEM 458 478
FT DOMAIN 479 501
FT ACT_SITE 93 93
FT ACT_SITE 289 289
FT CARBOHYD 153 153
FT CARBOHYD 172 172
FT CARBOHYD 223 223
FT CARBOHYD 354 354
FT CARBOHYD 501 AA; 55763 MW; 377CEAC824ACEF05 CRC64;
SQ SEQUENCE 501 AA; 55763 MW; 377CEAC824ACEF05 CRC64;
Query Match 89.1%; Score 1673; DB 1; Length 501;
Best Local Similarity 95.3%; Pred. No. 1.4e-137;
Matches 322; Conservative 1; Mismatches 7; Indels 8; Gaps 2;
QY 3 PTI--YLOAHFTLCGWSSTYRDLKRGVVPYTOGKWEGLGTDLVSPHGNVTVRANI 60
DB 107 PFLHYRYQQL-----SSTYRDLKRGVVPYTOGKWEGLGTDLVSPHGNVTVRANI 160
QY 61 AATTSDEKFFINGSNWEGTLGLAYAEIARPDSPPEFFDSLVKQTHVNLFLSLQLCGAGF 120
DB 161 AATTSDEKFFINGSNWEGTLGLAYAEIARPDSPPEFFDSLVKQTHVNLFLSLQLCGAGF 220
QY 121 PLNQSEVLASVSGSMIGGIDSHSLTGSLSWYTPIRREWYEVIIYVRVEINGDGLKMDCKE 180
DB 221 PLNQSEVLASVSGSMIGGIDSHSLTGSLSWYTPIRREWYEVIIYVRVEINGDGLKMDCKE 280
QY 181 YNYDKSIVDSGTTNLRPLPKKVFEEAAVKSIIKAASSTKPDGFWLGEOLVCWQAGTTPWNI 240
DB 281 YNYDKSIVDSGTTNLRPLPKKVFEEAAVKSIIKAASSTKPDGFWLGEOLVCWQAGTTPWNI 340

QY 241 FPVISLYLMGEVNTQSFRTIIPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVINE 300
DB 341 FPVISLYLMGEVNTQSFRTIIPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVINE 400
QY 301 GFVVVDFDRARKRIGFAVSACHVHDEFRTAAVEGPFCHL 338
DB 401 GFVVVDFDRARKRIGFAVSACHVHDEFRTAAVEGPFVTL 438
RESULT 2
BACE_MOUSE
ID BACE_MOUSE STANDARD; PRT: 501 AA.
AC P56818;
DT 30-MAY-2000 (Rel. 39, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE BETA-SECRETASE PRECURSOR (EC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)
DE (BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL
DE PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC PROTEASE 2)
DE (MEMAPsin-2).
GN BACE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20002972; PubMed=10531052;
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
RA Blere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,
RA Treanor J., Rogers G., Citron M.;
RA "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
RT the transmembrane aspartic protease BACE";
RN Science 286:735-741(1999).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057170; PubMed=10591213;
RA Van R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
RA Tomasselli A.G., Farodi L.A., Heinrikson R.L., Gurney M.E.;
RA "Membrane-anchored aspartyl protease with Alzheimer's disease
RT beta-secretase activity.";
RL Nature 402:533-537(1999).
CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP.
CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AF190726; AAF04143.1; -
CC EMBL; AF200346; AAF17082.1; -
CC MGD; MGI:1346542; BACE.
CC INTERPRO; IPR001461; -
CC PFAM; PF00026; asp; 3.
CC PRINTS; PR00792; PEPsin.
CC PROSITE; PS00141; ASP_PROTEASE; 1.
KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;

KW Signal.
 FT PROPEP 1 21 POTENTIAL.
 FT CHAIN 22 45 POTENTIAL.
 FT DOMAIN 46 501 BETA-SECRETASE.
 FT TRANSMEM 22 457 EXTRACELLULAR (POTENTIAL).
 FT ACT_SITE 458 478 POTENTIAL.
 FT ACT_SITE 479 501 CYTOPLASMIC (POTENTIAL).
 FT ACT_SITE 93 93 BY SIMILARITY.
 FT ACT_SITE 289 289 BY SIMILARITY.
 FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 6 6 R -> H (IN REF. 2).
 FT CONFLICT 81 87 RQPLTK -> VGSPPT (IN REF. 2).
 SQ SEQUENCE 501 AA; 55881 MW; B2FD346419376ACF CRC64;

Query Match 88.4%; Score 1661; DB 1; Length 501;
 Best Local Similarity 94.3%; Pred. No. 1.5e-136;
 Matches 316; Conservative 5; Mismatches 6; Indels 8; Gaps 2;

QY 3 PFI--YLQAHFTLCGWSSTYRDLRGVVPYVYQGWEGELGTDLSVPHGPNVTVRANI 60
 DB 107 PFLHRYQRQL-----SSTYRDLRGVVPYVYQGWEGELGTDLSVPHGPNVTVRANI 160

QY 61 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVQTHVPNLSLQLCGAGF 120
 DB 161 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVQTHVPNLSLQLCGAGF 220

QY 121 PLNQSEVLASVGSMTIIGDHSLSYTGSLWYTPIRREWYVEIIVRVEINGDLKMDCKE 180
 DB 221 PLNQTEALASVGSMTIIGDHSLSYTGSLWYTPIRREWYVEIIVRVEINGDLKMDCKE 280

QY 181 YNYDKSIVDSGTTNLRPKKVEAAVKSIAASSTKPEPDGFWLGEQLVCWQAGTTPWNI 240
 DB 281 YNYDKSIVDSGTTNLRPKKVEAAVKSIAASSTKPEPDGFWLGEQLVCWQAGTTPWNI 340

QY 241 FVVISLYLMGEVNTQSFRTILPQOYLPRVEDVATSDDDCYKFAISQSSTGTVMGAVIME 300
 DB 341 FVVISLYLMGEVNTQSFRTILPQOYLPRVEDVATSDDDCYKFAISQSSTGTVMGAVIME 400

QY 301 GFYVVDRAKRIGFAVSACHVHDEFRTAAVEGPF 335
 DB 401 GFYVVDRAKRIGFAVSACHVHDEFRTAAVEGPF 435

RESULT 3
 BACE_RAT STANDARD; PRT; 501 AA.
 AC P56819;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE BETA-SECRETASE PRECURSOR (RC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)
 DE (BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL
 DE PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC PROTEASE 2)
 DE (MEMAPSN-2).
 GN BACE.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20002972; PubMed-10531052;
 RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
 RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
 RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
 RA Biere A.L., Curran E., Burgess T., Louis J.C., Collins F.,
 RA Treanor J., Rogers G., Citron M.;
 RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
 RT the transmembrane aspartic protease BACE.";
 RL Science 286:735-741(1999).

CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP.
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
 CC FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF190727; AAF04144.1; -.
 CC INTERPRO: IPR001461; -.
 CC PFAM: PF00026; asp; 3.
 CC PRINTS: PR00792; PEPsin.
 CC PROSITE: PS00141; ASP_PROTEASE; 1.
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;
 KW Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 45 POTENTIAL.
 FT CHAIN 46 501 BETA-SECRETASE.
 FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 458 478 POTENTIAL.
 FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).
 FT ACT_SITE 93 93 BY SIMILARITY.
 FT ACT_SITE 289 289 BY SIMILARITY.
 FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 501 AA; 55806 MW; 24B445BC8BE87DE3 CRC64;

Query Match 88.1%; Score 1655; DB 1; Length 501;
 Best Local Similarity 94.0%; Pred. No. 5e-136;
 Matches 315; Conservative 5; Mismatches 7; Indels 8; Gaps 2;

QY 3 PFI--YLQAHFTLCGWSSTYRDLRGVVPYVYQGWEGELGTDLSVPHGPNVTVRANI 60
 DB 107 PFLHRYQRQL-----SSTYRDLRGVVPYVYQGWEGELGTDLSVPHGPNVTVRANI 160

QY 61 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVQTHVPNLSLQLCGAGF 120
 DB 161 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVQTHVPNLSLQLCGAGF 220

QY 121 PLNQSEVLASVGSMTIIGDHSLSYTGSLWYTPIRREWYVEIIVRVEINGDLKMDCKE 180
 DB 221 PLNQTEALASVGSMTIIGDHSLSYTGSLWYTPIRREWYVEIIVRVEINGDLKMDCKE 280

QY 181 YNYDKSIVDSGTTNLRPKKVEAAVKSIAASSTKPEPDGFWLGEQLVCWQAGTTPWNI 240
 DB 281 YNYDKSIVDSGTTNLRPKKVEAAVKSIAASSTKPEPDGFWLGEQLVCWQAGTTPWNI 340

QY 241 FVVISLYLMGEVNTQSFRTILPQOYLPRVEDVATSDDDCYKFAISQSSTGTVMGAVIME 300
 DB 341 FVVISLYLMGEVNTQSFRTILPQOYLPRVEDVATSDDDCYKFAISQSSTGTVMGAVIME 400

QY 301 GFYVVDRAKRIGFAVSACHVHDEFRTAAVEGPF 335
 DB 401 GFYVVDRAKRIGFAVSACHVHDEFRTAAVEGPF 435

RESULT 4
 BAE2_HUMAN STANDARD; PRT; 518 AA.
 ID BAE2_HUMAN

DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT PEP4_CHICK (EC 3.4.23.1)
 DE PEP4_CHICK (EC 3.4.23.1)
 OS Gadus morhua (Atlantic cod)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidae; Gadidae;
 OC Gadus.
 [1]
 RN SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.
 RP TISSUE-STOMACH;
 RC Karlsten S., Hough E., Olsen R.L.;
 RA "Structure and proposed amino-acid sequence of a pepsin from Atlantic
 RT cod (Gadus morhua).";
 RL Acta Crystallogr. D 54:32-46(1998).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 DR PDB; 1AM5; 24-DEC-97.
 DR INTERPRO: IPR001461; -
 DR INTERPRO: IPR001969; -
 DR PFAM; PF00026; asp; 1.
 DR PRINTS; PR00792; PEP4.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
 KW 3D-structure.
 FT ACT_SITE 32 32 BY SIMILARITY.
 FT ACT_SITE 214 214 BY SIMILARITY.
 FT DISULFID 45 50 BY SIMILARITY.
 FT DISULFID 206 209 BY SIMILARITY.
 FT DISULFID 247 280 BY SIMILARITY.
 SQ SEQUENCE 324 AA; 34014 MW; EE3A6097B6941DD7 CRC64;

Query Match 13.8%; Score 259; DB 1; Length 324;
 Best Local Similarity 28.7%; Pred. No. 3.3e-15;
 Matches 90; Conservative 48; Mismatches 114; Indels 62; Gaps 14;

QY 18 SSTYRDLRKGVVYPTQGWEGELGTLVSIPIHG--PNVTVRANIAAITESDKFFINGSN 75
 DB 61 SSTYVETGKIVDLTYGTGNGRILGQDTSVGGSDPNQELG---ESQTEPGFQA-AAP 116
 QY 76 WEGILGLAYAEIARPDSPPEFFDSLVKQTHV-PNLSQLQCGAGPLNQLSEVLASVGS 134
 DB 117 FDGILGLAYPSIAAA--GAVPVFDNMGSSQLVEKDLFSYLSGGG--ANGSEVM----- 166
 QY 135 MITGGIDHSYLTSLAWTPTRREWYVEIIVRVEINGDLKMD-CREYNVDKSIDSGTT 193
 DB 167 --LGGVDNSHTGSIHWIPVTAEKYQWALDGTIVNGQTACBGC-----QALVDTGTS 218
 QY 194 NLRLPKVFEAAVKSIAASSTKFFPDGFWLGEQLVCWQAGTTPWNIFFVISLYLMEV 253
 DB 219 KIVAPVSALANINKIGASEN-----QGEMWGN--CASVQSLPDITF-----T 259
 QY 254 NQSFRTILPQOYLRPVEDVATSDCCYKFAISOSSTGT-----VMGAVIMEGFYV 305
 DB 260 INGKQPLPSPSAIEGQAFCTSD-----GLGSGVSPNTSELWIFGDVFLRNYTI 310
 QY 306 FDRARKEIGFAVSA 319
 DB 311 YDRTNKVKGFAPAA 324

RESULT 6
 ID PEP4_CHICK STANDARD; PRT; 367 AA.
 AC P00793;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PEP4_CHICK A PRECURSOR (EC 3.4.23.1).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=84004412; PubMed=6617663;
 RA Baudys M., Kostka V.;
 RT "Covalent structure of chicken pepsinogen.";
 RL Eur. J. Biochem. 136:89-99(1983).
 CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
 CC ALSO CLEAVED TO SOME EXTENT.
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 DR PIR; A00984; PECH.
 DR HSP; P00794; 3CMS.
 DR MEROPS; A01.001; -
 DR INTERPRO: IPR001461; -
 DR INTERPRO: IPR001969; -
 DR PFAM; PF00026; asp; 1.
 DR PRINTS; PR00792; PEP4.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Stomach; Zymogen;
 KW Glycoprotein; Gastric juice.
 FT PROPEP 1 42 ACTIVATION PEPTIDE.
 FT CHAIN 43 367 PEPSIN A.
 FT ACT_SITE 77 77
 FT ACT_SITE 260 260
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .).
 FT DISULFID 90 95
 FT DISULFID 251 255
 FT DISULFID 290 323
 SQ SEQUENCE 367 AA; 40431 MW; OC547E7FD8F5B341 CRC64;

Query Match 13.5%; Score 253.5; DB 1; Length 367;
 Best Local Similarity 23.6%; Pred. No. 1.2e-14;
 Matches 80; Conservative 62; Mismatches 114; Indels 83; Gaps 13;

QY 2 VPFTYLOAHETLCSGW-----SSTYRDLRKGVVYPTQGWEGELGTLVSIPIHGPNV 54
 DB 85 VPSIYCKS--SACSNHRRFDPSSSTVSTNETVYIAYGTGSMGILGYDTVAV---SSI 139
 QY 55 TVRANIAAITSOK-FFINGSNWEGILGLAYAEIARPDSPPEFFDSLVKQTHV-PNLS 112
 DB 140 DVQNIQFGLSETPEGSFFCYCNFDGILGLAAPSIS--SSGATPVFDNMMSOHLVAQDLFS 197
 QY 113 LQLCGAGPLNQSEVLASVGSMTIIGDHSYLTSLAWTPTRREWYVEIIVRVEINGQ 172
 DB 198 VYLSKDG-----ETGSFVLFGGIDPNYTTKGIYVWPLSAETWQITMDRVTYGNK 247
 QY 173 DLK--MDCKEYNDKSIDSGTTNLRLPKVFEAAVKSIAASSTK-----KFPDGF 223
 DB 248 YVACFFTC-----QALVDTGTSLLVMPQAGNRIIKDLGVSSDGEISCDISKLPD--- 298
 QY 224 LGEOLVCWQAGTTPWNIFFVISLYLMEVNTNQSPRITLPQOYLRPVEDVATSDCCYK 283
 DB 299 -----VTFHINGHA-----FTLPASAYVLNEDGSCML 325
 QY 284 AISQSSTGT-----VMGAVIMEGFYVDFRARKRIGFA 316
 DB 326 GFENMGTPTELGEQWILGDVFIREFYVIFDRANKVGLS 364

RESULT 7
 ID PEP4_MACFU STANDARD; PRT; 388 AA.
 AC P27678;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PEP4_MACFU A PRECURSOR (EC 3.4.23.1) (PEPSIN I/II).
 GN PGA.
 OS Macaca fuscata fuscata (Japanese macaque).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.
RC TISSUE-GASTRIC MUCOSA;
RX MEDLINE=92037645; PubMed=1935977;
RA Kageyama T., Tanabe K., Koiwai O.;
RT "Development-dependent expression of isozymogens of monkey
RT pepsinogens and structural differences between them.";
RL Eur. J. Biochem. 202:205-215(1991).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY, ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPsin GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -!- MISCELLANEOUS: EACH PEPsin GENES IS CONVERTED TO CORRESPONDING
CC PEPsin AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA
CC ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC
CC CLEAVAGE VIA AN INTERMEDIATE FORM(S).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X59753; CAA42425.1; -
CC PIR; S16065; S16065.
CC PIR; S19682; S19682.
CC HSP; P00790; IPSO.
CC MEROPS; A01.001; -
CC INTERPRO; IPR001461; -
CC INTERPRO; IPR001969; -
CC PFAM; PF00026; asp; 1.
CC PRINTS; PR00792; PEPsin.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
CC Hydrolyase; Aspartyl protease; Digestion; Stomach; Gastric juice;
CC Zymogen; Multigene family; Signal; Glycoprotein.
CC SIGNAL 1 15 BY SIMILARITY.
CC PROPEP 16 38 ACTIVATION PEPTIDE.
CC CHAIN 39 62 ACTIVATION PEPTIDE.
CC FT ACT_SITE 94 94 PEPsin A-4.
CC FT ACT_SITE 277 277 BY SIMILARITY.
CC FT DISULFID 107 112 BY SIMILARITY.
CC FT DISULFID 268 272 BY SIMILARITY.
CC FT DISULFID 311 344 BY SIMILARITY.
CC FT CARBOHYD 88 88 N-LINKED (GLCNAC...).
CC SEQUENCE 388 AA; 41955 MW; A2923AB1F7FCDEB9 CRC64;

CC Query Match 13.4%; Score 251; DB 1; Length 388;
CC Best Local Similarity 28.7%; Pred. No. 2.1e-14;
CC Matches 89; Conservative 50; Mismatches 113; Indels 58; Gaps 13;

QY 18 SSTYRDLKRGVVPYTOCKWEGELGTLVSIPLHGPVTVVRANIAAITESSDK-FFINGSNW 76
DB 123 SSTYRATSKTVSITVGTGSMGILGYDIVK---GGISDTNQIFGLSETPGFFLYFAPF 179
QY 77 EGILGLAYAEIARPDSPPEPFDLSLVKQTHV-PNLFSLQLCAGPPLNQSEVLASVGGSM 135
DB 180 DGLGLAYPSIS--SSGATPVFDNWNQRLVSQDLFSVLSAD-----DQS-----GSV 227
QY 136 ITGGIDHSLSYTGSLWYTPRRWYVEVTVIRVEINGQDI--KMCKEYNDKSIYDSGTT 193
DB 228 IFGGIDSSYITGSLNWPVSVVEGYWQISVDSITMKGKTIACAKGC-----QAIVDTGTS 281

QY 194 NLRLPKKVFEEAAVKSIAAKSTKEKPDGFWLGEOLV-CWQAGTTPWNIFFVVISLYLMGEV 252
DB 282 LTGTPTSPIANIQSDIGASENSD-----GEMVVSICAISSLPDIIVF----- 322
QY 253 TNQSFRTILPQOY-LRPVEDVATSQDDCYK-----FAISOSSTGTVMGAVIMEGFYVVF 306
DB 323 -----TINGVOYPLPPSAYILOSGSCTSGFGMDVPTESGELWILGDVFIQYFIVF 375
QY 307 DRARKRIGFA 316
DB 376 DRANNQVGLA 385

RESULT 8
PEPsin_RABIT STANDARD; PRT; 387 AA.
ID PEPsin_RABIT
AC P28712;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPsin II-1 PRECURSOR (EC 3.4.23.1) (PEPsin A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=91009127; PubMed=2129536;
CC Kageyama T., Tanabe K., Koiwai O.;
CC "Structure and development of rabbit pepsinogens. Stage-specific
CC zymogens, nucleotide sequences of cDNAs, molecular evolution, and
CC gene expression during development";
CC J. Biol. Chem. 265:17031-17038(1990).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY, ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -!- DEVELOPMENTAL STAGE: PEPsin GENES IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPsin GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC PIR; B38302; B38302.
CC HSP; P00791; IPSA.
CC MEROPS; A01.001; -
CC INTERPRO; IPR001461; -
CC INTERPRO; IPR001969; -
CC PFAM; PF00026; asp; 1.
CC PRINTS; PR00792; PEPsin.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
CC Hydrolyase; Aspartyl protease; Digestion; Stomach; Gastric juice;
CC Zymogen; Signal; Phosphorylation; Multigene family.
CC SIGNAL 1 15 ACTIVATION PEPTIDE.
CC PROPEP 16 59 ACTIVATION PEPTIDE.
CC CHAIN 60 387 PEPsin II-1.
CC MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
CC FT ACT_SITE 93 93 BY SIMILARITY.
CC FT ACT_SITE 276 276 BY SIMILARITY.
CC FT DISULFID 106 111 BY SIMILARITY.
CC FT DISULFID 267 271 BY SIMILARITY.
CC FT DISULFID 310 343 BY SIMILARITY.
CC SEQUENCE 387 AA; 42070 MW; A6EC48F715541A48 CRC64;

CC Query Match 12.7%; Score 239; DB 1; Length 387;
CC Best Local Similarity 26.3%; Pred. No. 2.3e-13;
CC Matches 83; Conservative 56; Mismatches 114; Indels 62; Gaps 14;

QY 18 SSTYRDLKRGVVPYTOCKWEGELGTLVSIPLHGPVTVVRANIAAITESSDK-FFINGSNW 74
DB 122 SSTPQATSETLSITGYTGSMGILGYDIVK---GNIEDTNQIFGLSKTEPGITFLV--A 176
QY 75 NWEGILGLAYAEIARPDSPPEPFDLSLVKQTHV-PNLFSLQLCAGPPLNQSEVLASVGG 133


```
Db 177 PFDGILGLAYPSISASDAT--PVFDNNMNEGLVSEDLFSVYLSNG-----EKGS 224
QY 134 SMIIIGGDHSLYTGSLWYTPIRREYIEVLIVRVEINGDLKW--DCKEYNYDKSIVDSG 191
Db 225 VMVFGGIDSSYYTGSLNNVPSVHEGYWQITMDSITTINGETIACADSC-----QAVYDTG 278
QY 192 TTNLRPLPKVFEAAVKSIAKASTKPPDGFWMGEQLV-CWQAGTTPWNFFPVSISLYLMG 250
Db 279 TSLLAGPTSIAISQSYGASKNL-----LGENIISCAIDSLPDIVF----- 321
QY 251 EVTNQSPRITLPOOYLRLPVED-VATSQDDC-----YKPAISQSSTGT--VMGAVIMEGFYV 304
Db 322 -----TINNVPYPLPASAVILKDDCLSGFGDMNLDTSYGELWILGDVFIROYFT 372
QY 305 VFDRARKRIGFAVSA 319
Db 373 VFDRANNQVGLAAAA 387

RESULT 9
PEP3_RABIT
ID PEP3_RABIT STANDARD; PRT; 387 AA.
AC P27822;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPSIN III PRECURSOR (EC 3.4.23.1) (PEPSIN A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -I- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOCENS AT LATE POSTNATAL STAGE.
CC -I- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M59237; AAA85370.1; -
CC PIR; E38302; E38302.
CC HSP; P00791; 1PSA.
CC MEROPS; A01.001; -.
CC INTERPRO; IPR001461; -.
CC INTERPRO; IPR001969; -.
CC PFAM; PF00026; asp; 1.
CC PRINTS; P00792; PEPSIN.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59 ACTIVATION PEPTIDE.
FT CHAIN 60 387 PEPSIN III.
FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
```

```
FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT DISULFID 106 111 BY SIMILARITY.
FT DISULFID 267 271 BY SIMILARITY.
FT DISULFID 310 343 BY SIMILARITY.
SQ SEQUENCE 387 AA; 41969 MW; 15A59AC81F36F9EF CRC64;

Query Match 12.6%; Score 237.5; DB 1; Length 387;
Best Local Similarity 26.0%; Pred. No. 3.1e-13; Indels 65; Gaps 15;
Matches 87; Conservative 62; Mismatches 121;

QY 2 VPRTYIQAHTFLCSGW-----SSYRDLRKGVYVPYTGQKWEGELGTDLVSIHPGPNV 54
Db 101 VPSVYCSS--AACSVHQNFPEDSSTFQATSESLSIYGTGSMTGFLGYDKVKV--GNI 155
QY 55 TVRANIAAITYESDK-FFINGSNWEGILGLAYAEATRPDDSPPEPFDLSLVKQTHV-PNLFS 112
Db 156 EDTNQIFGLSEPGSFLYAPFDGILGLAYPSISSSDAT--PVFDNNMNEGLVSEDLFS 213
QY 113 LQLGAGGFPLNQSEVLASVGSMITGIDHSLYTGSLWYTPIRREYIEVLIVRVEINGQ 172
Db 214 VYLSDD-----ESGVVMFGGIDSSYYTGSNNWVPVSYEGYWIQLDITMDGE 263
QY 173 DLKM--DCKEYNYDKSIVDSGTTNLRPLPKVFEAAVKSIAKASTKPPDGFWMGEQLVC 230
Db 264 TIACADSC-----QAIVDTGTSLLAGP-----TSAISNQSYIGASENSD---GEMIVS 309
QY 231 WQAGTTPWNFFPVSILYMGVEVNTQSPRITLPOOYLRLPVEDVATSQDDC-----YKPAIS 286
Db 310 CS-----SMYSLPNIV-----FTINGVQVVPASAYILEEDDACISGFEGMNL 352
QY 287 QSSTGT--VMGAVIMEGFYVVDPRARKRIGFAVSA 319
Db 353 DTYTGELWILGDVFIROYFTVFDRANNQVGLAAAA 387

RESULT 10
PEP2_MACFU
ID PEP2_MACFU STANDARD; PRT; 388 AA.
AC P27677;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPSIN A-2/A-3 PRECURSOR (EC 3.4.23.1) (PEPSIN III-2/III-1).
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 16-70.
RC TISSUE-GASTRIC MUCOSA;
RX MEDLINE=92037645; PubMed=1935977;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Development-dependent expression of isozymogens of monkey
RT pepsinogens and structural differences between them.";
RL Eur. J. Biochem. 202:205-215(1991).
CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -I- DEVELOPMENTAL STAGE: PEP A-2 IS PREDOMINANT AT THE 4-MONTH STAGE.
CC PEP A-3 IS PREDOMINANT AT FETAL STAGES.
CC -I- PTM: PEP A-2 IS PHOSPHORYLATED, BUT NOT PEP A-3.
CC -I- PTM: EACH PEPSINOGEN IS CONVERTED TO CORRESPONDING PEPSIN AT PH
CC 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA ACTIVATION
CC SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC CLEAVAGE
CC VIA AN INTERMEDIATE FORM(S).
CC -I- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -I- CAUTION: IT IS NOT KNOWN IF THIS IS PEP A-2 OR PEP A-3.
```


Db 355 SGELWILGDVFIQRYTVDFRANKVGLA 383

RESULT 12

CATE_RAT STANDARD; PRT; 398 AA.

AC P16228; Q63701;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE CATHEPSIN E PRECURSOR (EC 3.4.23.34).
 GN CRSE.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SPLEEN;
 RX MEDLINE=96004785; PubMed=7574663;
 RA Okamoto K., Yu H., Misumi Y., Ikehara Y., Yamamoto K.;
 RT "Isolation and sequencing of two cDNA clones encoding rat spleen
 RT cathepsin E and analysis of the activation of purified procathepsin
 RT E.";
 RL Arch. Biochem. Biophys. 322:103-111(1995).
 RN [2]
 RP SEQUENCE OF 59-110.
 RX MEDLINE=90147750; PubMed=2105725;
 RA Yonezawa S., Takahashi T., Ichinose M., Miki K., Tanaka J., Gasa S.;
 RT "Structural studies of rat cathepsin E: amino-terminal structure and
 RT carbohydrate units of mature enzyme.";
 RL Biochem. Biophys. Res. Commun. 166:1032-1038(1990).
 CC -!- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
 CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
 CC -!- CATALYTIC ACTIVITY: SIMILAR TO CATHEPSIN D, BUT SLIGHTLY BROADER
 CC SPECIFICITY.
 CC -!- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 DR EMBL; D45187; BAA08128.1; -;
 DR EMBL; D38104; BAA07285.1; -;
 DR PIR; A34657; A34657.
 DR PIR; B34643; B34643.
 DR PIR; C34643; C34643.
 DR HSSP; P00794; 3CMS.
 DR MEROPS; A01.010; -;
 DR INTERPRO; IPR001461; -;
 DR INTERPRO; IPR001969; -;
 DR PFAM; PF00026; asp. 1.
 DR PRINTS; P00792; PEPsin
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 DR Hydrolyase: Aspartyl protease; Glycoprotein; Zymogen; Signal;
 KW Alternative splicing.
 FT SIGNAL 1 19 POTENTIAL.
 FT PROPEP 20 58 ACTIVATION PEPTIDE.
 FT CHAIN 59 398 CATHEPSIN E.
 FT MOD_RES 20 20 PYROLIDONE CARBOXYLIC ACID
 (BY SIMILARITY).
 FT ACT_SITE 98 98 BY SIMILARITY.
 FT ACT_SITE 283 283 BY SIMILARITY.
 FT DISULFID 62 62 INTERCHAIN (PROBABLE).
 FT DISULFID 111 116 BY SIMILARITY.
 FT DISULFID 274 278 BY SIMILARITY.
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT VARSPPLIC 312 344 MISSING (IN SHORT ISOFORM).

FT VARIANT 114 114 P -> S.
 FT CONFLICT 79 E -> N (IN REF. 2).
 FT CONFLICT 84 TV -> SR (IN REF. 2).
 SQ SEQUENCE 398 AA; 43021 MW; 25F123E67C46EB5F CRC64;

Query Match 12.4%; Score 233; DB 1; Length 398;

Best Local Similarity 25.4%; Pred. No. 7.8e-13;
 Matches 84; Conservative 59; Mismatches 130; Indels 58; Gaps 14;

QY 2 VPFYI-----LQAHFTLCGMSSTYRDLRKGVVYPTQKWEGLGTDLVSPHGPNTV 56

Db 106 VPSVYCTSPACKAHPVFPSPQSTMEVGNHFSIOYGTGSLTGIIIGADQVSV-EGLTVEG 164

QY 57 RANIAITSDREFFINGSNWEGLTGLAYAEIARPDSPPEFDSLVKQTHVP-NLFSQL 115

Db 165 QQFGESVKEPGQTFVN-AEFDGLGLGYPSLA--VGVVTPVFDNMAQNLVALPHEVYL 221

QY 116 CGAGFPLNQSEVLASVGGSMIIIGDHSLYTGLSWYTPIRREMYEVIIVRVEINGQDLK 175

Db 222 -----SSDPQGGSGSELTTGGYDPSHFSGSLNIPVTQGYMQIALDGIQVG--DTV 271

QY 176 MCKEYNYDKSIVDSGTTNLRPLPKVFEAAVKSIAKASSTEFPPDGFGLGQLVCWQAGT 235

Db 272 MFCSE--GCQAIVDGTGTLTGTPPK----KIKQLQEA-----IGA 305

QY 236 TPNWIFFVSLYLMGEVNTQSFRI-----TILPQQVLRPVEDVATSDDC---YKFAISQ 287

Db 306 TPDGGEYVDCATLNMVNTFLINGSVYTLSTAVILP-DLVDMQFGSGFGQLDIO 363

QY 288 SSTGT--VMGAVIMEGVVYVDFRARRKIGFA 316

Db 364 PPAGPLWILGDVFIQRYTVDFRANKVGLA 394

RESULT 13

PEPA_HUMAN STANDARD; PRT; 388 AA.

AC P00790;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PEPsin A PRECURSOR (EC 3.4.23.1).
 GN PGA3 AND PGA4 AND PGAS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83161158; PubMed=6300126;
 RA Sogawa K., Fujii-Kuriyama Y., Mizukami Y., Ichihara Y., Takahashi K.;
 RT "Primary structure of human pepsinogen gene.";
 RL J. Biol. Chem. 258:5306-5311(1983).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOZYME 5).
 RC TISSUE=PLACENTA;
 RX MEDLINE=89233110; PubMed=2714789;
 RA Evers M.P.J., Zelle B., Bebelman J.P., van Beusechem V., Kraakman L.,
 RA Hoffer M.J.V., Pronk J.C., Mager W.H., Planta R.J., Eriksson A.W.,
 RA Frants R.R.;
 RT "Nucleotide sequence comparison of five human pepsinogen A (PGA)
 RT genes: evolution of the PGA multigene family.";
 RL Genomics 4:232-239(1989).
 RN [3]
 RP PARTIAL SEQUENCE OF 1-28.
 RX MEDLINE=86059312; PubMed=2415509;
 RA Ichihara Y., Sogawa K., Takahashi K.;
 RT "Isolation of human, swine, and rat prepepsinogens and calf
 RT preprochymosin, and determination of the primary structures of their
 RL NH2-terminal signal sequences.";
 RL J. Biochem. 98:483-492(1985).
 RN [4]
 RP SEQUENCE OF 16-100 (ISOZYME 2; 3; 3A; 4 AND 5).

RX MEDLINE-90130402; PubMed-2515193;
 RA Athauda S.B.P., Tanji M., Kageyama T., Takahashi K.;
 RT "A comparative study on the NH2-terminal amino acid sequences and
 some other properties of six isozymic forms of human pepsinogens and
 pepsins."; J. Biochem. 106:920-927(1989).
 RL [5]
 RN SEQUENCE OF 16-68 (ISOZYMES 3 AND 5).
 RP MEDLINE-89065108; PubMed-3197840;
 RX Foltmann B.;
 RA "Activation of human pepsinogens."; FEBS Lett. 241:69-72(1988).
 RT [6]
 RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RP MEDLINE-95392399; PubMed-7663352;
 RA Fujinaga M., Chernaia M.M., Tarasova N.I., Mosimann S.C.,
 RA James M.N.G.;
 RT "Crystal structure of human pepsin and its complex with pepstatin."; Protein Sci. 4:960-972(1995).
 RL CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE ALSO CLEAVED TO SOME EXTENT.
 CC CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
 CC CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC CC -----
 DR EMBL; J00287; AAA98529.1; JOINED.
 DR EMBL; J00279; AAA98529.1; JOINED.
 DR EMBL; J00280; AAA98529.1; JOINED.
 DR EMBL; J00281; AAA98529.1; JOINED.
 DR EMBL; J00282; AAA98529.1; JOINED.
 DR EMBL; J00283; AAA98529.1; JOINED.
 DR EMBL; J00284; AAA98529.1; JOINED.
 DR EMBL; J00285; AAA98529.1; JOINED.
 DR EMBL; J00286; AAA98529.1; JOINED.
 DR EMBL; M26032; AAA60061.1; JOINED.
 DR EMBL; M26025; AAA60061.1; JOINED.
 DR EMBL; M26026; AAA60061.1; JOINED.
 DR EMBL; M26027; AAA60061.1; JOINED.
 DR EMBL; M26028; AAA60061.1; JOINED.
 DR EMBL; M26029; AAA60061.1; JOINED.
 DR EMBL; M26030; AAA60061.1; JOINED.
 DR EMBL; M26031; AAA60061.1; JOINED.
 DR PIR; A00980; PEHU.
 DR PIR; A30142; B30142.
 DR PIR; B30142; B30142.
 DR PIR; A22434; A22434.
 DR PIR; PX0023; PX0023.
 DR PIR; PX0024; PX0024.
 DR PIR; PX0025; PX0025.
 DR PIR; PX0026; PX0026.
 DR PIR; PX0027; PX0027.
 DR PIR; S02663; S02663.
 DR PIR; S02664; S02664.
 DR PDB; 1PSN; 20-APR-95.
 DR PDB; 1PSO; 20-APR-95.
 DR MEROPS; A01.001; -.
 DR MIM; 169700; -.
 DR MIM; 169710; -.
 DR MIM; 169720; -.
 DR MIM; 169730; -.
 DR INTERPRO; IPR001461; -.
 DR INTERPRO; IPR001969; -.
 DR PFAM; PF00026; asp. 1.
 DR PRINTS; PR00792; PEPSIN.

DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
 KW Zymogen; Signal; Phosphorylation; 3D-structure; Polymorphism.
 FT SIGNAL 1 15
 FT PROPEP 16 62 ACTIVATION PEPTIDE.
 FT CHAIN 63 388 PEPSIN A.
 FT MOD_RES 130 130 PHOSPHORYLATION (POTENTIAL).
 FT ACT_SITE 94 94
 FT ACT_SITE 277 277
 FT DISULFID 107 112 BY SIMILARITY.
 FT DISULFID 268 272 BY SIMILARITY.
 FT DISULFID 311 344 BY SIMILARITY.
 FT VARIANT 28 28 L -> F (IN ISOZYME 5).
 FT VARIANT 58 58 E -> K (IN ISOZYME 3A, 4 AND 5).
 FT VARIANT 92 92 V -> L (IN ISOZYME 4 AND 5).
 FT VARIANT 222 222 Q -> K.
 FT VARIANT 265 265 A -> T.
 FT VARIANT 353 353 L -> V.
 FT VARIANT 376 376 D -> E.
 FT SEQUENCE 388 AA; 41977 MW; C9CB89BA08F4D788 CRC64;
 Query Match 12.3%; Score 231.5; DB 1; Length 388;
 Best Local Similarity 27.0%; Pred. No. 1e-12;
 Matches 90; Conservative 56; Mismatches 120; Indels 67; Gaps 15;
 QY 2 VPFTYLOA-----HFTLCGWSSTYRDLKGVYVYPTQCKWEGELGTDLVSIHPGNVTV 56
 DB 102 VPSVYCSLACTNHNRFNPEDSSYQSTSETVSYTGTSMTGILGYDTQV---GGISD 158
 QY 57 RANAAITESDK-PFINGSNWEGILGLAYAEIARPDSPPEPFDSLVRKQTHV-PNLFSIQ 114
 DB 159 TNQIFGLSETEPGSFLYAPDGLGLAYPSIS--SGATPVFDNIWNOGLVSDQLFSVY 216
 QY 115 LCGAGFLPNQSEVLASVGGSMIIGIDHSLYTSLWYTPIRREWYEVYIVRVEINGDL 174
 DB 217 LSAD-----DQS-----GSVIFGIDSSYYTGSLSNWVPTVEGYWQITVDSITMNGEAI 266
 QY 175 KMDCKEYNDKSIIVDSGTTNLRPKKVFEEAAVSIKAASSTKEKPDGFWLGEQLV-CWQA 233
 DB 267 A--CAE--GCQAIYDTGTSLITGPTSPANTQSDIGASENSD-----GDMVYSCSAI 314
 QY 234 GTTPWNIFPVVISLVMGEVINSFRITLPOOY-LRPVEDVATSDDDCYKFAIS----- 286
 DB 315 SSLPDIVF-----TINGVQVPPVPSAYILQSEGSC-----ISGFOGMN 352
 QY 287 ---QSTGTVMGAVIMEGFYVDFRARKRIGFA 316
 DB 353 LPTEGELWILGVDVFIQYFTVFDRAANQVGLA 385
 RESULT 14
 ID PEPA_MACMU STANDARD; PRT; 388 AA.
 AC P11489;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PEPSIN A PRECURSOR (EC 3.4.23.1).
 GN PGA.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
 OC Cercopitheciinae; Macaca.
 RN [1]
 RP SEQUENCE FROM N.A.

357 SGELWILGDVFRQYFTVFDNRANNVGLA 385

DB

RESULT 15

PEPE_CHICK STANDARD; PRT; 383 AA.

AC P16476;

AC 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE EMBRYONIC PEPSINOGEN PRECURSOR (EC 3.4.23.-).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OC [1]

SEQUENCE FROM N.A.

RP MEDLINE=88227903; PubMed=3131317;

RA Hayashi K., Agata K., Mochii M., Yasugi S., Eguchi G., Mizuno T.;

RA "Molecular cloning and the nucleotide sequence of cDNA for embryonic

RT chicken pepsinogen: phylogenetic relationship with prothymosin.";

RL J. Biochem. 103:290-296(1988).

CC -!- DEVELOPMENTAL STAGE: SPECIFICALLY SECRETED DURING THE EMBRYONIC

CC PERIOD IN THE CHICKEN PROVENTRICULUS (GLANDULAR STOMACH).

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE

CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; D00215; BAA00153.1; -.

CC PIR; A41443; A41443.

DR HSSP; P00794; 3CMS.

DR MEROPS; A01.028; -.

DR INTERPRO; IPR001461; -.

DR INTERPRO; IPR001969; -.

DR PFAM; PF00026; asp.1.

DR PRINTS; P00792; PEPsin.

DR PROSITE; PS00141; ASP_PROTEASE; 2.

KW Hydrolase; Aspartyl protease; Digestion; Signal; Glycoprotein.

FT SIGNAL 1 16

FT CHAIN 17 383

FT ACT_SITE 94 94

FT ACT_SITE 276 276

FT DISULFID 107 112

FT DISULFID 267 271

FT DISULFID 310 344

FT CARBOHYD 132 132

FT CARBOHYD 204 204

FT CARBOHYD 309 309

FT CARBOHYD 350 350

FT VARIANT 51 51

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

FT T -> S.

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT T -> S.

FT SEQUENCE 3

```
Db 219 -REFGSMVFGGIDSYFTGSINWIPVSYQGYWQISMDSIIVNKQEIACSSGC----- 271
QY 185 KSI VDSGTTNLR LPKKYFEAAVKSIAASSTERFPDGF LGEQLVCHQAGTTPWNI EPVI 244
Db 272 QAIIDTGTSLVAGPASDINDIQSAVGANQNT-----YGEYSV-----NCSHIL 314
QY 245 SL-----YIMGEVTNQSFRITILPQQYLRPVEDVA-----TSQDDCYKFAISQSSTGTVMGA 296
Db 315 AMPDVVFVIGGI-----QY--PVPALAYTEQNGGQTCMSSSFQNSSADLWILGD 360
QY 297 VIMEGFYVFDRAKRIGFA 316
Db 361 VFIRVYYSIEDRANRVGLA 380
```

Search completed: March 24, 2001, 13:24:20
Job time: 166 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:21:29 ; Search time 47.63 Seconds

(without alignments)
251.985 Million cell updates/sec

Title: US-09-215-435-167

Perfect score: 1878

Sequence: 1 MWPFIYLAHFTLCGSWSST.....EGPFCHLGHRLWLQHSR 351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36:*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1878	100.0	351	20 Y35918	Extended human sec
2	1673	89.1	453	21 Y88438	Modified human asp
3	1673	89.1	459	21 Y88439	Modified human asp
4	1668	88.8	425	21 Y88437	Human Asp2 amino a
5	1668	88.8	433	21 Y88433	Human-pro-Asp-2(a)
6	1668	88.8	446	21 Y88431	T7-caspase-human-p
7	1668	88.8	459	21 Y88432	T7-caspase-human-p
8	1668	88.8	501	21 Y88425	Human aspartyl pro
9	1667	88.8	501	19 W59807	Amino acid sequenc
10	1661	88.4	501	21 Y88427	Murine aspartyl pr
11	1656	88.2	790	19 W59808	Partial amino acid
12	1525.5	81.2	476	21 Y88426	Human aspartyl pro

13	925	49.3	518	19 W61362	Aspartic proteinas
14	925	49.3	518	20 Y41714	Human PRO852 prote
15	925	49.3	518	20 Y22239	Human CSP56, aspar
16	925	49.3	518	20 Y13799	Human aspartyl pro
17	925	49.3	518	21 Y88424	Human aspartyl pro
18	292	15.5	53	20 Y11427	Human 5' EST secre
19	228	12.1	381	13 R20730	Prochymosin (prore
20	224	11.9	412	16 R74207	Human death associ
21	224	11.9	412	19 W71369	Death associated p
22	224	11.9	412	20 Y06478	Human tumour-asso
23	223	11.9	365	4 P30603	Sequence encoded b
24	223	11.9	365	11 R05080	Sequence of calf p
25	223	11.9	375	5 P40078	Sequence encoded b
26	223	11.9	380	3 P20038	pre-prorennin-A pr
27	223	11.9	381	5 P40559	Sequence of a poly
28	222	11.8	381	5 P40218	Sequence of rennin
29	219	11.7	365	10 P94144	Prochymosin. P94
30	219	11.7	458	10 P94376	BamHI/Sali insert
31	219	11.7	545	20 Y33830	Oleolin-spacer-Met
32	218	11.6	381	4 P30086	Sequence encoded b
33	216	11.5	388	20 Y32058	Cat pregnancy asso
34	215	11.4	381	4 P30446	Sequence encoded b
35	214	11.4	379	4 P30013	Sequence encoded b
36	203	10.8	375	20 Y32055	Bovine pregnancy a
37	199	10.6	391	20 Y32056	Bovine pregnancy a
38	197.5	10.5	376	20 Y32054	Bovine pregnancy a
39	197	10.5	450	10 P94370	Sequence encoded b
40	196.5	10.5	392	20 Y32057	Bovine pregnancy a
41	196	10.4	376	20 Y32035	Bovine pregnancy a
42	193	10.3	380	20 Y32041	Bovine pregnancy a
43	193	10.3	387	20 Y32052	Bovine pregnancy a
44	185.5	9.9	381	20 Y32046	Bovine pregnancy a
45	185	9.9	450	21 Y57041	Plasmodium vivax p

ALIGNMENTS

RESULT	1
Y35918	
ID	Y35918 standard; Protein; 351 AA.
XX	
AC	Y35918;
XX	
XX	
DT	13-SEP-1999 (first entry)
DE	Extended human secreted protein sequence, SEQ ID NO. 167.
XX	
KW	Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW	cellular differentiation; immune system regulator; anti-inflammatory;
KW	haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW	reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
XX	genetic disease.
OS	Homo sapiens.
XX	
PN	WO9931236-A2.
XX	
PD	24-JUN-1999.
XX	
PF	17-DEC-1998; 98WO-IB02122.
XX	
PR	10-AUG-1998; 98US-0096116.
PR	17-DEC-1997; 97US-0069957.
PR	09-FEB-1998; 98US-0074121.
PR	13-APR-1998; 98US-0081563.
XX	
PA	(GEST) GENSET.
XX	
PI	Bougueleret L, Duclert A, Dumas Milne Edwards J;
DR	WPI: 1999-385906/32.
DR	N-PSDB; X97602.

XX PT New isolated human secreted proteins
 XX PS Claim 9; Page 198-199; 516pp; English.
 XX CC This sequence is encoded by an extended human secreted protein coding
 CC sequence of the invention. The secreted proteins can be used in treating
 CC or controlling a variety of human conditions. The secreted proteins may
 CC act as cytokines or may affect cellular proliferation or differentiation
 CC or may act as immune system regulators, haematopoiesis regulators, tissue
 CC growth regulators, regulators of reproductive hormones or cell movement
 CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
 CC tumour inhibition activity. The DNAs can be used in forensic procedures
 CC to identify individuals or in diagnostic procedures to identify
 CC individuals having genetic diseases resulting from abnormal expression of
 CC the genes corresponding to the extended cDNAs. They are also useful for
 CC constructing a high resolution map of the human chromosomes. They can
 CC also be used for gene therapy to control or treat genetic diseases.
 XX CC Sequence 351 AA;
 XX SQ

Query Match 100.0%; Score 1878; DB 20; Length 351;
 Best Local Similarity 100.0%; Pred. No. 9.9e-197;
 Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYPFIYQLAHFTLCGWSSTYRDLRKGVVYPTQCKWEGELGTDLSVPHGPNVTVRANI 60
 DB 1 mvpfiylqahftlcgwsstyrdlrkgvvpytgqkwegeigtdivsphgpnvtvrani 60
 QY 61 AATFSDKFFINGNWEGILGLAYAEIARPDSPFPFDSLVKQTHVPLNLSLQICGAGF 120
 DB 61 aaatesdkffingsnwegilglayaeiarppdspeffdslvkqthvplnslqlcagaf 120
 QY 121 PLNQSEVLASVGGSMIIGDHSLYTGSLSWYTPPIREWYVEIIVRVEINGDLKMDCKE 180
 DB 121 plnqsevlasvggsmliigdhsltygslwytppirrewyveilvrveingdldkmdcke 180
 QY 181 YNYDKSIVDSGTTNLRPKKVFEEAVKSIKAASSTKPEKPDGFWLGEQLVCWQAGTTPWNI 240
 DB 181 ynydksivdsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttwni 240
 QY 241 FPVISLYLMGEVTNQSFRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIME 300
 DB 241 fpvislylmgevtngsfrilitpqoylrpvedvatsddcykfaissqstgtvmgavime 300
 QY 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVGGPCHLGHRLWQHSTDR 351
 DB 301 gfyvvfdrarkrigfavsachvhdefrtaavegppchlhghrlwqhstdr 351

RESULT 2
 Y88438
 ID Y88438 standard; Protein; 453 AA.
 XX AC Y88438;
 XX DT 03-AUG-2000 (first entry)
 XX DE Modified human aspartyl protease 2 (Asp2) amino acid sequence.
 KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 KW Alzheimer's disease; beta secretase site.
 XX OS Homo sapiens.
 XX PN WO200017369-A2.
 XX PD 30-MAR-2000.
 XX PF 23-SEP-1999; 99WO-0520881.
 XX PX 24-SEP-1998; 98US-0101594.

XX PA (PHAA) PHARMACIA & UPJOHN CO.
 XX PI Gurney ME, Bienkowski MJ, Heintz RL, Parodi LA, Yan R;
 XX DR N-PSDB; A15688.
 XX WI: 2000-303209/26.
 XX PT New enzyme designated human aspartase useful in research into
 PT Alzheimer's disease is capable of cleaving amyloid protein precursor at
 PT the beta secretase site to produce amyloid beta peptide -
 XX Example 10; Page 169-172; 183pp; English.
 XX CC This sequence represents a modified human aspartyl protease 2 (Asp2)
 CC amino acid sequence. Asp2 encoded by this sequence has the C-terminal
 CC transmembrane domain deleted. The invention relates to a protease
 CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
 CC precursor protein (APP). The protease contains a sequence encoding the
 CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
 CC 100-300 amino acids. When mutated the APP gene causes an autosomal
 CC dominant form of Alzheimer's disease. APP localises to the cell surface
 CC membrane and have a single C-terminal transmembrane domain. Proteolytic
 CC processing of APP produces the amyloid beta protein, which is possibly
 CC very important in Alzheimer's disease. The invention includes a
 CC nucleotide sequence encoding the protease, a vector containing the
 CC screening for inhibitors of beta secretase activity are also given in the
 CC invention. The human aspartase protein and nucleotide sequences and the
 CC methods for identifying inhibitors of the protease, are useful in the
 CC treatment of and research in to Alzheimer's disease.
 XX CC Sequence 453 AA;
 XX SQ

Query Match 89.18; Score 1673; DB 21; Length 453;
 Best Local Similarity 95.3%; Pred. No. 3.7e-174;
 Matches 322; Conservative 1; Mismatches 7; Indels 8; Gaps 2;

QY 3 PFI--YLQAHFTLCGWSSTYRDLRKGVVYPTQCKWEGELGTDLSVPHGPNVTVRANI 60
 DB 107 pfihryygrql-----sstydrirkgvvpytgqkwegeigtdivsphgpnvtvrani 160
 QY 61 AATFSDKFFINGNWEGILGLAYAEIARPDSPFPFDSLVKQTHVPLNLSLQICGAGF 120
 DB 161 aaatesdkffingsnwegilglayaeiarppdspeffdslvkqthvplnslqlcagaf 220
 QY 121 PLNQSEVLASVGGSMIIGDHSLYTGSLSWYTPPIREWYVEIIVRVEINGDLKMDCKE 180
 DB 221 plnqsevlasvggsmliigdhsltygslwytppirrewyveilvrveingdldkmdcke 280
 QY 181 YNYDKSIVDSGTTNLRPKKVFEEAVKSIKAASSTKPEKPDGFWLGEQLVCWQAGTTPWNI 240
 DB 281 ynydksivdsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttwni 340
 QY 241 FPVISLYLMGEVTNQSFRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIME 300
 DB 341 fpvislylmgevtngsfrilitpqoylrpvedvatsddcykfaissqstgtvmgavime 400
 QY 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVGGPCHL 338
 DB 401 gfyvvfdrarkrigfavsachvhdefrtaavegppfvtl 438

RESULT 3
 Y88439
 ID Y88439 standard; Protein; 459 AA.
 XX AC Y88439;
 XX DT 03-AUG-2000 (first entry)
 XX PX Modified human aspartyl protease 2 (Asp2) amino acid sequence.

```
XX Asparyl protease; aspartase; amyloid precursor protein; APP; Asp 2;  
KW Alzheimer's disease; beta secretase site.  
XX  
OS Homo sapiens.  
XX  
PN WO200017369-A2.  
XX  
PD 30-MAR-2000.  
XX  
PF 23-SEP-1999; 99WO-US20881.  
XX  
PR 24-SEP-1998; 98US-0101594.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.  
PI Gurney ME, Bienkowski MJ, Heinrichson RL, Parodi LA, Yan R;  
XX  
XX WPI; 2000-303209/26.  
DR N-PSDB; A15689.  
XX  
PT New enzyme designated human aspartase useful in research into  
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at  
PT the beta secretase site to produce amyloid beta peptide -  
XX  
XX Example 10; Page 173-176; 183pp; English.  
XX  
XX This sequence represents a modified human aspartyl protease 2 (Asp2)  
CC amino acid sequence. Asp2 encoded by this sequence has the C-terminal  
CC transmembrane domain deleted. The invention relates to a protease  
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid  
CC precursor protein (APP). The protease contains a sequence encoding the  
CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by  
CC 100-300 amino acids. When mutated the APP gene causes an autosomal  
CC dominant form of Alzheimer's disease. APP localises to the cell surface  
CC membrane and have a single C-terminal transmembrane domain. Proteolytic  
CC processing of APP produces the amyloid beta protein, which is possibly  
CC very important in Alzheimer's disease. The invention includes a  
CC nucleotide sequence encoding the protease, a vector containing the  
CC nucleotide sequence, and a cell line comprising the vector. Methods for  
CC screening for inhibitors of beta secretase activity are also given in the  
CC invention. The human aspartase protein and nucleotide sequences and the  
CC methods for identifying inhibitors of the protease, are useful in the  
CC treatment of and research in to Alzheimer's disease.  
XX  
SQ Sequence 459 AA;  
  
Query Match 89.1%; Score 1673; DB 21; Length 459;  
Best Local Similarity 95.3%; Pred. No. 3.8e-174;  
Matches 322; Conservative 1; Mismatches 7; Indels 8; Gaps 2;  
  
QY 3 PFI--YLOAHFTLCGWSSTYRDLRKGVVYPYTGQKWEGLGTLVSIHPGPNVTVRANI 60  
Db 107 pfhryyyqrql-----sstyrdlrgvvyptqgkwegelgtldvsihpgpnvtvrani 160  
QY 61 AAITESDKFFINGSNWEGTLGLAYAEIARPDSPPEFFDSLKVQTHVPLNLSQLCGAGF 120  
Db 161 aaitesdkffingsnwegtlglayaeiarpddslpeffdsldskvqthvplnlsqicgagf 220  
QY 121 PLNQSEVLASVSGSMIIIGIDHSLTGSLWYTPIRREWYEVIIVRVEINGODLKMDCKE 180  
Db 221 plnqsevlavsgsmiigidhslytgslywtpirreyeviivrveingodlkmcdke 280  
QY 181 YNYKSIDVSGTNNLRPKKVFEEAVKSKAASSTKPPDGFGLGEQLVCWQAGTTPWNI 240  
Db 281 ynyksidvsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttwni 340  
QY 241 FVVISLYLMEVTNOSFRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIME 300  
Db 341 fpvislylmevtnosfrilitlpqoylrpvedvatsddcykfaissqstgtvmgavime 400  
QY 301 GFYVVFDRARRKRGFAVSACHVHDEFRTAAVEGPFCHL 338
```

```
Db 401 gfyvvfdrrarkrgfavsachvhdefrtaavegpfvtl 438  
RESULT 4  
Y88437  
ID Y88437 standard; Protein; 425 AA.  
XX  
AC Y88437;  
XX  
DT 03-AUG-2000 (first entry)  
XX  
DE Human Asp2 amino acid sequence containing proteolytic cleavage site.  
XX  
KW Asparyl protease; aspartase; amyloid precursor protein; APP; Asp 2;  
KW Alzheimer's disease; beta secretase site.  
XX  
OS Homo sapiens.  
XX  
PN WO200017369-A2.  
XX  
PD 30-MAR-2000.  
XX  
PF 23-SEP-1999; 99WO-US20881.  
XX  
PR 24-SEP-1998; 98US-0101594.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.  
PI Gurney ME, Bienkowski MJ, Heinrichson RL, Parodi LA, Yan R;  
XX  
XX WPI; 2000-303209/26.  
DR N-PSDB; A15677.  
XX  
XX New enzyme designated human aspartase useful in research into  
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at  
PT the beta secretase site to produce amyloid beta peptide -  
XX  
XX Example 9; Page 166-168; 183pp; English.  
XX  
XX This sequence represents a modified version of the human aspartase 2  
CC (Asp2) nucleotide sequence. The sequence is used in the bacterial  
CC expression of human Asp2L. The invention relates to a protease  
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid  
CC precursor protein (APP). The protease contains a sequence encoding the  
CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by  
CC 100-300 amino acids. When mutated the APP gene causes an autosomal  
CC dominant form of Alzheimer's disease. APP localises to the cell surface  
CC membrane and have a single C-terminal transmembrane domain. Proteolytic  
CC processing of APP produces the amyloid beta protein, which is possibly  
CC very important in Alzheimer's disease. The invention includes a  
CC nucleotide sequence encoding the protease, a vector containing the  
CC nucleotide sequence, and a cell line comprising the vector. Methods for  
CC screening for inhibitors of beta secretase activity are also given in the  
CC invention. The human aspartase protein and nucleotide sequences and the  
CC methods for identifying inhibitors of the protease, are useful in the  
CC treatment of and research in to Alzheimer's disease.  
XX  
SQ Sequence 425 AA;  
  
Query Match 88.8%; Score 1668; DB 21; Length 425;  
Best Local Similarity 95.0%; Pred. No. 1.2e-173;  
Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;  
  
QY 3 PFI--YLOAHFTLCGWSSTYRDLRKGVVYPYTGQKWEGLGTLVSIHPGPNVTVRANI 60  
Db 79 pfhryyyqrql-----sstyrdlrgvvyptqgkwegelgtldvsihpgpnvtvrani 132  
QY 61 AAITESDKFFINGSNWEGTLGLAYAEIARPDSPPEFFDSLKVQTHVPLNLSQLCGAGF 120  
Db 133 aaitesdkffingsnwegtlglayaeiarpddslpeffdsldskvqthvplnlsqicgagf 192
```

QY 121 PLNOSVLASVSGSMIIIGIDHSYTSIWTPIRREWYVEIIVRVEINGDLMCKE 180
 DB 193 plngsevlavsgsmiigldhslytgslytprirrewyveilvrveingdldmcke 252
 QY 181 YNYDKSIIVDSGTTNLRPKKVFEEAAVSIKAASTKFPDGFVLGEQLVCWQAGTTPWNI 240
 DB 253 ynydksivdsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagtpwni 312
 QY 241 FVISLYLMGEVNTNOSFRITILPQOYLPRPVEDVATSDDCYKFAISQSSTGTVMGAVIME 300
 DB 313 fpvislylmgevtngsfritilpqgylrpvedvatsqdcykfaissgstgtvmgavime 372
 QY 301 GFYVVDRAKRIGFAVSACHVHDEFRTAAVEGPFCHL 338
 DB 373 gfyvvdrrarkrigfavsachvhdefrtaavegpfvltl 410

RESULT 5
 Y88433
 ID Y88433 standard; Protein; 433 AA.
 XX
 AC Y88433;
 XX
 XX 03-AUG-2000 (first entry)
 XX
 DE Human-pro-Asp-2(a)-deltaTM amino acid sequence.
 XX
 KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 KW Alzheimer's disease; beta secretase site; human-pro-Asp-2(a)-deltaTM.
 XX
 OS Chimeric - Homo sapiens.
 OS Chimeric - Bacteriophage T7.
 XX
 PN WO200017369-A2.
 XX
 PD 30-MAR-2000.
 XX
 PF 23-SEP-1999; 99WO-US20881.
 XX
 PR 24-SEP-1998; 98US-0101594.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX
 PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
 XX
 XX WPI; 2000-303209/26.
 DR N-PSDB; A15670.
 XX
 XX New enzyme designated human aspartase useful in research into
 PT Alzheimer's disease is capable of cleaving amyloid protein precursor at
 PT the beta secretase site to produce amyloid beta peptide -
 XX
 XX Example 9; Fig 8; 183pp; English.
 PS
 CC This sequence represents a modified version of the human aspartase 2
 CC (Asp2) amino acid sequence. The sequence is used in the bacterial
 CC expression of human Asp2L. The invention relates to a protease
 CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
 CC precursor protein (APP). The protease contains a sequence encoding the
 CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
 CC 100-300 amino acids. When mutated the APP gene causes an autosomal
 CC dominant form of Alzheimer's disease. APP localises to the cell surface
 CC membrane and have a single C-terminal transmembrane domain. Proteolytic
 CC processing of APP produces the amyloid beta protein, which is possibly
 CC very important in Alzheimer's disease. The invention includes a
 CC nucleotide sequence encoding the protease, a vector containing the
 CC nucleotide sequence, and a cell line comprising the vector. Methods for
 CC screening for inhibitors of beta secretase activity are also given in the
 CC invention. The human aspartase protein and nucleotide sequences and the
 CC methods for identifying inhibitors of the protease, are useful in the
 CC treatment of and research in to Alzheimer's disease.
 XX
 XX Sequence 433 AA;

Query Match 88.8%; Score 1668; DB:21; Length 433;
 Best Local Similarity 95.0%; Pred. No. 1.2e-173;
 Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;
 QY 3 PFI--YLQAHFTLCGWSSTVRLRKGVYVYPTQGWEGELGTDLVSIPIHGPNVTVRANI 60
 DB 87 pfllhryyqrql-----sstyrdirksvvyptqgkwegelgtclvsipghpnvtrani 140
 QY 61 AAITESDKFFINGNWGILGLAYAEATARPDDSPPEFDSLVKQTHVPLNLSLQLCGAGF 120
 DB 141 aaaitesdkffingnwegilglayaeatarpddslepfdsivkqthvplnlslnlcagaf 200
 QY 121 PLNOSVLASVSGSMIIIGIDHSYTSIWTPIRREWYVEIIVRVEINGDLMCKE 180
 DB 201 plngsevlavsgsmiigldhslytgslytprirrewyveilvrveingdldmcke 260
 QY 181 YNYDKSIIVDSGTTNLRPKKVFEEAAVSIKAASTKFPDGFVLGEQLVCWQAGTTPWNI 240
 DB 261 ynydksivdsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagtpwni 320
 QY 241 FVISLYLMGEVNTNOSFRITILPQOYLPRPVEDVATSDDCYKFAISQSSTGTVMGAVIME 300
 DB 321 fpvislylmgevtngsfritilpqgylrpvedvatsqdcykfaissgstgtvmgavime 380
 QY 301 GFYVVDRAKRIGFAVSACHVHDEFRTAAVEGPFCHL 338
 DB 381 gfyvvdrrarkrigfavsachvhdefrtaavegpfvltl 418

RESULT 6
 Y88431
 ID Y88431 standard; Protein; 446 AA.
 XX
 AC Y88431;
 XX
 XX 03-AUG-2000 (first entry)
 XX
 DE T7-caspase-human-pro-Asp-2(a)-deltaTM amino acid sequence.
 XX
 KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 KW Alzheimer's disease; beta secretase site;
 KW T7-caspase-human-pro-Asp-2(a)-deltaTM.
 XX
 OS Chimeric - Homo sapiens.
 OS Chimeric - Bacteriophage T7.
 XX
 PN WO200017369-A2.
 XX
 PD 30-MAR-2000.
 XX
 PF 23-SEP-1999; 99WO-US20881.
 XX
 PR 24-SEP-1998; 98US-0101594.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX
 PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
 XX
 XX WPI; 2000-303209/26.
 DR N-PSDB; A15668.
 XX
 XX New enzyme designated human aspartase useful in research into
 PT Alzheimer's disease is capable of cleaving amyloid protein precursor at
 PT the beta secretase site to produce amyloid beta peptide -
 XX
 XX Example 9; Fig 6; 183pp; English.
 PS
 CC This sequence represents a modified version of the human aspartase 2
 CC (Asp2) amino acid sequence. The sequence is used in the bacterial
 CC expression of human Asp2L. The invention relates to a protease
 CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
 CC precursor protein (APP). The protease contains a sequence encoding the
 CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
 CC 100-300 amino acids. When mutated the APP gene causes an autosomal
 CC dominant form of Alzheimer's disease. APP localises to the cell surface
 CC membrane and have a single C-terminal transmembrane domain. Proteolytic
 CC processing of APP produces the amyloid beta protein, which is possibly
 CC very important in Alzheimer's disease. The invention includes a
 CC nucleotide sequence encoding the protease, a vector containing the
 CC nucleotide sequence, and a cell line comprising the vector. Methods for
 CC screening for inhibitors of beta secretase activity are also given in the
 CC invention. The human aspartase protein and nucleotide sequences and the
 CC methods for identifying inhibitors of the protease, are useful in the
 CC treatment of and research in to Alzheimer's disease.
 XX
 XX Sequence 433 AA;

CC precursor protein (APP). The protease contains a sequence encoding the
 CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
 CC 100-300 amino acids. When mutated the APP gene causes an autosomal
 CC dominant form of Alzheimer's disease. APP localises to the cell surface
 CC membrane and have a single C-terminal transmembrane domain. Proteolytic
 CC processing of APP produces the amyloid beta protein, which is possibly
 CC very important in Alzheimer's disease. The invention includes a
 CC nucleotide sequence encoding the protease, a vector containing the
 CC screening for inhibitors of beta secretase activity are also given in the
 CC invention. The human aspartase protein and nucleotide sequences and the
 CC methods for identifying inhibitors of the protease, are useful in the
 CC treatment of and research in to Alzheimer's disease.
 XX
 SQ Sequence 446 AA;

Query Match 88.8%; Score 1668; DB 21; Length 446;
 Best Local Similarity 95.0%; Pred. No. 1.3e-173;
 Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;
 Qy 3 PFI--YLOAHFTLCGWSSTYDLRKGVYVPYTOGKWEGLGTLVSIPIHGNVTVRANI 60
 Db 100 pflhryyqrql-----sstyrdirkgvvyptqgkwegelgtdivsphgpnvtvrani 153
 Qy 61 AAITESDRFFINGSNWEGILGLAYAEIARPDSPFPFDSLKQTHVPLNLSQLCGAGF 120
 Db 154 aaatesdkffingsnwegilglayaeiarpdsslepfdsivkqthvplnlsfhlhcgagf 213
 Qy 121 PLNQSEVLASVGGSMIIGSDHSLYTGSLWTPPIRREWYVEIIVRVEINGODLKMDCKE 180
 Db 214 plnqsevlasvggsmiigsdhsltygslwtpirrewyveilvrveingodlkmdccke 273
 Qy 181 YNYDKSIYDSTGTTNLRPKKVFEEAAVKSIAASSTKEKFPDGFGLGEQLVCMQAGTTPWNI 240
 Db 274 ynydksiydsdgttnlrpkpkvfeaaavksikaasstekfpdgfwlgeqlvcwaggttpwni 333
 Qy 241 FPVISLYLMGEVTNOSFRITILPQQYLPRVEDVATSDQDCKYFAISQSSTGVMGAVIME 300
 Db 334 fpvislylmgevtngsfritilpqyilrpvedvatsqddcykfaissqstgtvmgavime 393
 Qy 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVGPFCHL 338
 Db 394 gfyvvfdrarkrigfavsachvhdefrtaavegpfvtl 431

RESULT 7
 Y88432
 ID Y88432 standard; Protein; 459 AA.
 XX
 AC Y88432;
 XX
 DT 03-AUG-2000 (first entry)
 XX
 DE T7-caspase-human-pro-Asp-2(a)-deltaTM amino acid sequence.
 XX
 KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 KW Alzheimer's disease; beta secretase site;
 KW T7-caspase-human-pro-Asp-2(a)-deltaTM.

XX Chimeric - Homo sapiens.
 OS Chimeric - Bacteriophage t7.
 XX
 PN WO200017369-A2.
 XX
 PD 30-MAR-2000.
 XX
 PF 23-SEP-1999; 99WO-US20881.
 XX
 PR 24-SEP-1998; 98US-0101594.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX

PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
 DR WPI; 2000-303209/26.
 DR N-PSDB; A15669.
 XX
 PT New enzyme designated human aspartase useful in research into
 PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
 PT the beta secretase site to produce amyloid beta peptide -
 XX
 PS Example 9; Fig 7; 183pp; English.
 XX
 CC This sequence represents a modified version of the human aspartase 2
 CC (Asp2) nucleotide sequence. The sequence is used in the bacterial
 CC expression of human Asp2L. The invention relates to a protease
 CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
 CC precursor protein (APP). The protease contains a sequence encoding the
 CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
 CC 100-300 amino acids. When mutated the APP gene causes an autosomal
 CC dominant form of Alzheimer's disease. APP localises to the cell surface
 CC membrane and have a single C-terminal transmembrane domain. Proteolytic
 CC processing of APP produces the amyloid beta protein, which is possibly
 CC very important in Alzheimer's disease. The invention includes a
 CC nucleotide sequence encoding the protease, a vector containing the
 CC screening for inhibitors of beta secretase activity are also given in the
 CC invention. The human aspartase protein and nucleotide sequences and the
 CC methods for identifying inhibitors of the protease, are useful in the
 CC treatment of and research in to Alzheimer's disease.
 XX
 SQ Sequence 459 AA;

Query Match 88.8%; Score 1668; DB 21; Length 459;
 Best Local Similarity 95.0%; Pred. No. 1.3e-173;
 Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;
 Qy 3 PFI--YLOAHFTLCGWSSTYDLRKGVYVPYTOGKWEGLGTLVSIPIHGNVTVRANI 60
 Db 113 pflhryyqrql-----sstyrdirkgvvyptqgkwegelgtdivsphgpnvtvrani 166
 Qy 61 AAITESDRFFINGSNWEGILGLAYAEIARPDSPFPFDSLKQTHVPLNLSQLCGAGF 120
 Db 167 aaatesdkffingsnwegilglayaeiarpdsslepfdsivkqthvplnlsfhlhcgagf 226
 Qy 121 PLNQSEVLASVGGSMIIGSDHSLYTGSLWTPPIRREWYVEIIVRVEINGODLKMDCKE 180
 Db 227 plnqsevlasvggsmiigsdhsltygslwtpirrewyveilvrveingodlkmdccke 286
 Qy 181 YNYDKSIYDSTGTTNLRPKKVFEEAAVKSIAASSTKEKFPDGFGLGEQLVCMQAGTTPWNI 240
 Db 287 ynydksiydsdgttnlrpkpkvfeaaavksikaasstekfpdgfwlgeqlvcwaggttpwni 346
 Qy 241 FPVISLYLMGEVTNOSFRITILPQQYLPRVEDVATSDQDCKYFAISQSSTGVMGAVIME 300
 Db 347 fpvislylmgevtngsfritilpqyilrpvedvatsqddcykfaissqstgtvmgavime 406
 Qy 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVGPFCHL 338
 Db 407 gfyvvfdrarkrigfavsachvhdefrtaavegpfvtl 444

RESULT 8
 Y88425
 ID Y88425 standard; Protein; 501 AA.
 XX
 AC Y88425;
 XX
 DT 03-AUG-2000 (first entry)
 XX
 DE Human aspartyl protease 2 (a) (Asp2) amino acid sequence.
 XX
 KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 KW Alzheimer's disease; beta secretase site.

```

XX OS Homo sapiens.
XX PN WO200017369-A2.
XX PD 30-MAR-2000.
XX PF 23-SEP-1999; 99WO-US20881.
XX PR 24-SEP-1998; 98US-0101594.
XX PA (PHAA ) PHARMACIA & UPJOHN CO.
XX PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
XX DR WPI; 2000-303209/26.
XX DR N-PSDB; A15662.
XX PT New enzyme designated human aspartase useful in research into
XX PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
XX PT the beta secretase site to produce amyloid beta peptide
XX PS Claim 48; Fig 2; 183pp; English.
XX CC This sequence represents the human aspartyl protease 2 (Asp2) amino acid
XX CC sequence. The invention relates to a protease (e.g. Asp2) capable of
XX CC cleaving the beta secretase site of amyloid precursor protein (APP). The
XX CC protease contains a sequence encoding the amino acid sequence DIG and a
XX CC sequence encoding DSG or DRG separated by 100-300 amino acids. When
XX CC mutated the APP gene causes an autosomal dominant form of Alzheimer's
XX CC disease. APP localises to the cell surface membrane and have a single
XX CC C-terminal transmembrane domain. Proteolytic processing of APP produces
XX CC the amyloid beta protein, which is possibly very important in Alzheimer's
XX CC disease. The invention includes a nucleotide sequence encoding the
XX CC protease, a vector containing the nucleotide sequence, and a cell line
XX CC comprising the vector. Methods for screening for inhibitors of beta
XX CC secretase activity are also given in the invention. The human aspartase
XX CC protein and nucleotide sequences and the methods for identifying
XX CC inhibitors of the protease, are useful in the treatment of and research
XX CC in to Alzheimer's disease.
XX SQ Sequence 501 AA;
Query Match 88.8%; Score 1668; DB 21; Length 501;
Best Local Similarity 95.0%; Pred. No. 1.5e-173;
Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;
QY 3 PFI--YLOAHFTCLSGWSSTYRDLRKGVVYPYTOGKWEGLGTLVSIPIHGNVTVRANI 60
DB 107 pfhryyqrql-----sstyrdlrgvypytgqkwegelgtdivsihphgnvtvrani 160
QY 61 AAITESDKFFINGSNWEGILGAYAEIARPDSPPEFFDSLKVQTHVPLNLSQLCGAGF 120
DB 161 aaatesdkffingsnwegilglayaeiarpdpslepfdsivkqthvplnlsqcgagf 220
QY 121 PLNQSEVLASVGGSMIIGIDHSILYTGSLWYTPIRREWYEVIIIVRVEINGDLKMDCKE 180
DB 221 plnqsevlasvggsmiigidhslytgslywtpirrewyevilvrveingdldkmdcke 280
QY 181 YNYDKSIVDSGTTNLRPKKVFEEAAVKSIAASSTKEKPDGFWLGEQLVCWQAGTTPWNI 240
DB 281 ynydksivdsgttnlrpkkvfeavksikaasstekfpdgfwlgeqlvcwqagttpwni 340
QY 241 FPVISLYLMGEVTNQSFRITILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIME 300
DB 341 fpvislylmgevtngsfrilitlppqylrpvedvatsqddcykfaissqstgtvmgavime 400
QY 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVGPFCCHL 338
DB 401 gfyvvfdrarkrigrfavsachvhdefrtaavegpfvtl 438

```

```

RESULT 9
W59807 W59807 standard; Protein; 501 AA.
XX AC W59807;
XX DT 26-OCT-1998 (first entry)
XX DE Amino acid sequence of human ASP2 (aspartic protease 2).
XX KW Human; ASP2; aspartic protease 2; agonist; antagonist; immunospecific;
XX KW antibody; inhibition; Alzheimer's disease; cancer; proteinase;
XX KW prohormone processing.
XX OS Homo sapiens.
XX PN EP855444-A2.
XX PD 29-JUL-1998.
XX PF 27-JAN-1998; 98EP-0300573.
XX PR 28-JAN-1997; 97GB-0001684.
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX PI Chapman CG, Murphy K, Powell DJ, Smith TS;
XX DR WPI; 1998-389809/34.
XX DR N-PSDB; V41696.
XX PT New nucleic acid encoding human aspartic protease 2 - used to treat,
XX PT prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone
XX PT processing
XX PS Claim 1; Page 7; 26pp; English.
XX CC This is the amino acid sequence of the human ASP2 (aspartic protease
XX CC family), used in the method of the invention. Agonists and
XX CC antagonists for ASP2 immunospecific antibodies are used to treat
XX CC conditions requiring increased or decreased activity or expression of
XX CC ASP2 respectively. ASP2 is used to treat and diagnose e.g.
XX CC Alzheimer's disease, cancer and prohormone processing and ASP2 or a
XX CC fragment can be used to induce an immune response against the above
XX CC conditions.
XX SQ Sequence 501 AA;
Query Match 88.8%; Score 1667; DB 19; Length 501;
Best Local Similarity 95.0%; Pred. No. 2e-173;
Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;
QY 3 PFI--YLOAHFTCLSGWSSTYRDLRKGVVYPYTOGKWEGLGTLVSIPIHGNVTVRANI 60
DB 107 pfhryyqrql-----sstyrdlrgvypytgqkwegelgtdivsihphgnvtvrani 160
QY 61 AAITESDKFFINGSNWEGILGAYAEIARPDSPPEFFDSLKVQTHVPLNLSQLCGAGF 120
DB 161 aaatesdkffingsnwegilglayaeiarpdpslepfdsivkqthvplnlsqcgagf 220
QY 121 PLNQSEVLASVGGSMIIGIDHSILYTGSLWYTPIRREWYEVIIIVRVEINGDLKMDCKE 180
DB 221 plnqsevlasvggsmiigidhslytgslywtpirrewyevilvrveingdldkmdcke 280
QY 181 YNYDKSIVDSGTTNLRPKKVFEEAAVKSIAASSTKEKPDGFWLGEQLVCWQAGTTPWNI 240
DB 281 ynydksivdsgttnlrpkkvfeavksikaasstekfpdgfwlgeqlvcwqagttpwni 340
QY 241 FPVISLYLMGEVTNQSFRITILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIME 300
DB 341 fpvislylmgevtngsfrilitlppqylrpvedvatsqddcykfaissqstgtvmgavime 400

```


FT Misc-difference 770 /label= unknown
FT /note= "encoded by GNC"
FT Misc-difference 776
FT /label= unknown
FT /note= "encoded by NTT"
FT Misc-difference 779
FT /label= unknown
FT /note= "encoded by GNG"
FT Misc-difference 780
FT /label= unknown
FT /note= "encoded by NCC"
FT Misc-difference 781
FT /label= unknown
FT /note= "encoded by NCT"
FT Misc-difference 782
FT /label= unknown
FT /note= "encoded by TNT"
FT Misc-difference 783
FT /label= unknown
FT /note= "encoded by ATN"
FT Misc-difference 784
FT /label= unknown
FT /note= "encoded by TNA"
FT Misc-difference 785
FT /label= unknown
FT /note= "encoded by TNC"
FT Misc-difference 790
FT /label= unknown
FT /note= "encoded by NNG"
FT
FT
XX EP855444-A2.
PN
XX
XX 29-JUL-1998.
XX
XX 27-JAN-1998; 98EP-0300573.
XX
XX 28-JAN-1997; 97GB-0001684.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX (SMIK) SMITHKLINE BEECHAM PLC.
XX
XX Chapman CG, Murphy K, Powell DJ, Smith TS;
XX
XX WPI; 1998-389809/34.
XX N-PSDB; V41697.
XX
XX New nucleic acid encoding human aspartic protease 2 - used to treat,
XX prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone
XX processing
XX
XX Disclosure; Pages 9; 26pp; English.
XX
XX This is the amino acid sequence of the partial human ASP2 (aspartic
XX protease 2), used in the method of the invention. Agonists and
XX antagonists for ASP2 immunospecific antibodies are used to treat
XX conditions requiring increased or decreased activity or expression
XX of ASP2 respectively. ASP2 is used to treat and diagnose e.g.
XX Alzheimer's disease, cancer and prohormone processing and ASP2 or a
XX fragment can be used to induce an immune response against the above
XX conditions.
XX
XX Sequence 790 AA;
XX
XX Query Match 88.2%; Score 1656; DB 19; Length 790;
XX Best Local Similarity 94.4%; Pred. No. 6.4e-172;
XX Matches 319; Conservative 1; Mismatches 10; Indels 8; Gaps 2;
XX
XX 3 PFT--VLOAHFTLCGWSSTYRDLKRGVVPYTGKWEGLCTDLVSIPIHGNVTVRANI 60
XX 11; 11
XX 50 pflhryyyrql-----sstyrdlrgkyepytgqkwegelgtdivsihghpntvrani 103

QY 61 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEPFDSLVKQTHVPNLFSLQLCGAGF 120
DB 104 aaitesdkffingsnwegilglayaeiarppdslepffdsllvqkthvnpnlfslqlcgagf 163
QY 121 PLNQSEVLASVGSGMIIGGDHSLYTGSLWYTPIRRWYVEYIIVRVEINGQDLKMDCKE 180
DB 164 plnqsevlasvgsgmiiggdhsllytgslwlytpirrewyveviivrvveingqdlkmdcke 223
QY 181 YNYDKSIVDSGTTNLRPKKVEEAANKSIRAAASSTERFDPGFWLGEOLVCMQAGTTPWNI 240
DB 224 ynydksivdsgettnlrlpkkveaavksikaasprekfgdgfwlgeglvcwqggttwnl 283
QY 241 FPVISLYLMGEVNTNQSFRTILPQQYLRPVEDVATSDDCYKFAISQSSSTGTVMGAVIME 300
DB 284 fpvislylmgevtngsfrtilpqgylrpvedvatsddcykfaigsstgtvmgavime 343
QY 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVEGPFCHL 338
DB 344 gfyvvfdrarkrigfavsachvhdefrtaavegpfvtl 381
RESULT 12
Y88426
ID Y88426 standard; Protein; 476 AA.
XX
XX Y88426;
XX
XX 03-AUG-2000 (first entry)
XX
XX Human aspartyl protease 2 (b) (ASP2) amino acid sequence.
XX
XX Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
XX Alzheimer's disease; beta secretase site.
XX
XX Homo sapiens.
XX
XX WO200017369-A2.
XX
XX 30-MAR-2000.
XX
XX 23-SEP-1999; 99WO-US20881.
XX
XX 24-SEP-1998; 98US-0101594.
XX
XX (PHAA) PHARMACIA & UPJOHN CO.
XX
XX Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
XX
XX WPI; 2000-303209/26.
XX N-PSDB; A15663.
XX
XX New enzyme designated human aspartase useful in research into
XX Alzheimer's Disease is capable of cleaving amyloid protein precursor at
XX the beta secretase site to produce amyloid beta peptide
XX
XX Claim 51; Fig 3; 183pp; English.
XX
XX This sequence represents the human aspartyl protease 2 (ASP2) amino acid
XX sequence. The invention relates to a protease (e.g. Asp2) capable of
XX cleaving the beta secretase site of amyloid precursor protein (APP). The
XX protease contains a sequence encoding the amino acid sequence DTG and a
XX sequence encoding DSG or DTG separated by 100-300 amino acids. When
XX mutated the APP gene causes an autosomal dominant form of Alzheimer's
XX disease. APP localises to the cell surface membrane and have a single
XX C-terminal transmembrane domain. Proteolytic processing of APP produces
XX the amyloid beta protein, which is possibly very important in Alzheimer's
XX disease. The invention includes a nucleotide sequence encoding the
XX protease, a vector containing the nucleotide sequence, and a cell line
XX comprising the vector. Methods for screening for inhibitors of beta
XX secretase activity are also given in the invention. The human aspartase
XX protein and nucleotide sequences and the methods for identifying
XX inhibitors of the protease, are useful in the treatment of and research
XX in to Alzheimer's disease.


```
xx SQ Sequence 476 AA;
Query Match 81.28; Score 1525.5; DB 21; Length 476;
Best Local Similarity 88.28; Pred. No. 5.2e-158;
Matches 298; Conservative 1; Mismatches 6; Indels 33; Gaps 3;
Qy 3 PFI--YLQAHFTLCGWSSTYRDLRKGVVYPTQGWKEGELGTDLVSLPHGPNVTVRANI 60
Db 107 pflhryyqrql-----sstyrdlrkgvvyptqgkwegelgtldvslphgpnvtrani 160
Qy 61 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEPFDLSLVKQTHVNPVLFSLQCGAGF 120
Db 161 aaitesdkffingsnwegilglayaeiar-----lcgagf 195
Qy 121 PLNQSEVLASVGGSMIIGGIDHSLYTGSLSWTPTRREHYEVIIVRVEINGQDLKMDCKE 180
Db 196 plnqsevlasvggsmiiggidhslytgslywtpirrehyeviivrveingqdlkmdcke 255
Qy 181 YNYSKSIYDVGSTTNLRPKVFEAAVKSIAASTSEKPPDGFVWLGEOQLVCWQAGTTPWNI 240
Db 256 ynyksiydvgsttnlrpkvfeaaavksiaastsekipdgfwlgeqlvcwqagttwni 315
Qy 241 FVISLYLMGEVTVNQSPRITILPQOYLRPVEDVATSDQDCYKFAISQSSSTGTVMGAVIME 300
Db 316 fpvislylmgevtngsfritilpqylrpvedvatsqddcykfaissqstgtvmgavime 375
Qy 301 GFYVFDPRKRIGFAVSACHVHDEFRTAAVEGPFCHL 338
Db 376 gfyvfdprkrigfavsachvhdefrtaavegpfvtl 413
```

```
RESULT 13
W61362
ID W61362 standard; Protein; 518 AA.
XX
AC W61362;
XX
DT 25-SEP-1998 (first entry)
XX
DE Aspartic proteinase ASPI.
XX
KW ASPI; Aspartic proteinase; Alzheimer's disease; cancer; melanoma.
XX
OS Homo sapiens.
XX
PN EP848062-A2.
XX
PD 17-JUN-1998.
XX
PF 01-DEC-1997; 97EP-0309648.
XX
PR 14-DEC-1996; 96GB-0026022.
XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI Chapman CG, Evans JR, Powell DJ, Southan C;
XX
DR WPI; 1998-314477/28.
XX
DR N-PSDB; V27962.
XX
PT New isolated polynucleotide encodes Aspartic protease polypeptide -
PT used to diagnosis, treat and vaccinate against Alzheimer's disease,
PT cancer and melanoma
XX
PS Claim 11; Page 7; 19pp; English.
XX
CC The human ASPI protein is structurally related to other proteins of the
CC Aspartic proteinase family. ASP 1 polypeptides and polynucleotides can
CC be used to diagnosis, treat and vaccinate against Alzheimer's disease,
CC cancer and melanoma.
```

```
xx SQ Sequence 518 AA;
Query Match 49.38; Score 925; DB 19; Length 518;
Best Local Similarity 52.38; Pred. No. 2.5e-92;
Matches 174; Conservative 56; Mismatches 97; Indels 6; Gaps 3;
Qy 3 PFYIQAHTLCSGWSSTYRDLRKGVVYPTQGWKEGELGTDLVSLPHGPNVTVRANIAA 62
Db 122 phsyldtyfd--tersstyrskgfdvtvkytqgswtgfvgedlvtipkgnfslvniat 179
Qy 63 ITESDKFFINGSNWEGILGLAYAEIARPDSPPEPFDLSLVKQTHVNPVLFSLQCGAGPPL 122
Db 180 ifesenffilpgikwngilglayatlakpssslstffdsilvtqanlpnvfsmqmcgaglpv 239
Qy 123 NOSEVLASVGGSMIIGGIDHSLYTGSLSWTPTRRREHYEVIIVRVEINGQDLKMDCKEYN 182
Db 240 ags---gtngsgslvlggielpslykgdiwytpikeewyvgieilkieigqgslnldcreyn 296
Qy 183 YDKSIVDVGSTTNLRPKVFEAAVKSIAASTSEKPPDGFVWLGEOQLVCWQAGTTPWNIFF 242
Db 297 adkaivdvgsttnlrpkvfdavvearaslipedsfgfwgtsqlactwnsetpwsyfp 356
Qy 243 VISLYLMGEVTVNQSPRITILPQOYLRPVEDVATSDQDCYKFAISQSSSTGTVMGAVIMEGF 302
Db 357 kisiylrdensrsfritilpqyliqpmmgaglny-ecyrfgispsstnalvigatvmegf 415
Qy 303 YVFDPRKRIGFAVSACHVHDEFRTAAVEGPF 335
Db 416 yvifdradrkrigfavsachvhdefrtaavegsf 448
```

```
RESULT 14
Y41714
ID Y41714 standard; Protein; 518 AA.
XX
AC Y41714;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PR0852 protein sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
OS Homo sapiens.
XX
PN WO9946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
```



```

XX PF 24-DEC-1998; 98WO-US27608.
XX PR 31-DEC-1997; 97US-0070112.
XX PA (CHIR ) CHIRON CORP.
XX PI Gliese K, Xin H;
XX WPI; 1999-430248/36.
XX DR N-PSDB; X84708.
XX PT New polynucleotides associated with cancer metastasis
XX PS Claim 4; Page 78-80; 80pp; English.
XX CC This sequence represents a polypeptide of the invention, and is
CC an aspartyl-type protease, designated CSP56. The polynucleotides (PNS) of
CC the invention encode metastatic marker protein variants. The PNS and
CC polypeptides can be used as markers for cancer metastasis. The PNS can
CC be used for identifying metastatic tissue or metastatic potential of
CC a tissue, e.g. breast or colon tissue. They can also be used for
CC screening test compounds for the ability to suppress the metastatic
CC potential of a tumour. The products can be used for developing products
CC for the therapy of cancers, particularly breast or colon cancer.
XX SQ Sequence 518 AA;

Query Match 49.38; Score 925; DB 20; Length 518;
Best Local Similarity 52.3%; Pred. No. 2.5e-92;
Matches 174; Conservative 56; Mismatches 97; Indels 6; Gaps 3

QY 3 PFYILQAHFTLCGWSSTSYRDLRKGVYVPYTGKGWEGELGTDLVSIPHGPNVTVRANIAA 62
DB 122 phsyidytdf--terstyrskgfdvtykytqgswtgrfvgedlvtipkgnfslvniat 179
QY 63 ITESDKFFINGSNWEGILGLAYAEIARDDSPPEPFDLSLVKQTHVPNLFSLQCGAGPPL 122
DB 180 ifesenfflpkgwlgilgavatlatkpsssletffdsilvtqanipnvfsmqmcgaglpv 239
QY 123 NQSEVLASVGSMIIGGDHSLYGLSWLYTPIRRWYVEVIVRVEINGQDLKMDCKEYN 182
DB 240 ags---gtngsylvlgiepslykgdiwyttptikeesyyvqgieilkiei9gqslnldcreyn 296
QY 183 YDKSIVDSGTTNLRLPKPVFAAVKYSIKAASSTKFPDGFWMLEQOLVCWQAGTTWNIFP 242
DB 297 adkaivdsqgtllrlpqkfvdaavearaslipefsdgfwgsglacwtnsetpwsyfp 356
QY 243 VISLYLGEVNTNQFRITILPQOYLRPVEDVATSDDCYKFAISQSSGTGTVMGAVIMEGF 302
DB 357 kislyldensrsfritilpqylqipmmgaglny-ecyrfgispsstnalvigatvmegf 415
QY 303 YVWFDRARKRTGFAVSAACHVHDEFRTAAVEGPF 335
DB 416 yvifdrakrvgfaaspcaeiaagavseisgpf 448

```

Search completed: March 24, 2001, 13:22:25
Job time: 56 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:21:34 ; Search time 73.46 Seconds
(without alignments)
560.032 Million cell updates/sec

Title: US-09-215-435-167
Perfect score: 1878
Sequence: 1 MVPEIYLQAHFTLCGWSST.....EGPFCHLGRLWLQHSNDR 351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.virus.*
- 13: sp.vertebrate.*
- 14: sp.unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1745	92.9	396	4 Q9ULS1	Q9uls1 homo sapien
2	913	48.6	514	11 Q9JL18	Q9jl18 mus musculus
3	784.5	41.8	213	4 Q9P0D2	Q9p0d2 homo sapien
4	767.5	40.9	396	4 Q9NZL1	Q9nzl1 homo sapien
5	726	38.7	468	4 Q9NZL2	Q9nzl2 homo sapien
6	559.5	29.8	255	11 Q9R1P7	Q9rlp7 mus musculus
7	254	13.5	423	5 Q9VKP7	Q9vkp7 drosophila
8	253.5	13.5	382	13 Q9PRG9	Q9prg9 gallus gall
9	253.5	13.5	391	5 Q9VLP6	Q9vlp6 drosophila
10	237.5	12.6	372	5 Q9VLK3	Q9vln3 drosophila
11	235	12.5	380	6 Q28950	Q28950 sus scrofa
12	233	12.4	396	13 Q93428	Q93428 chionodraco
13	232	12.4	346	6 Q9TS27	Q9ts27 bos taurus
14	232	12.4	384	13 Q91322	Q91322 rana catesb
15	230.5	12.3	385	6 Q29080	Q29080 sus scrofa
16	226.5	12.1	376	13 Q9PUR8	Q9pur8 pleuronecte
17	225.5	12.0	387	6 Q9N2D4	Q9n2d4 callithrix
18	225	12.0	381	6 Q9N2D2	Q9n2d2 callithrix
19	221	11.8	379	11 Q9JUX1	Q9jjx1 rattus norv

20	220.5	11.7	378	13 Q9PUR9	Q9pur9 pleuronecte
21	219	11.7	387	11 Q9JKE6	Q9jke6 mus musculus
22	216	11.5	388	6 Q46524	Q46524 felis silve
23	215	11.4	399	13 Q93458	Q93458 podarcis si
24	212.5	11.3	380	6 Q9TTX0	Q9ttx0 capra hircu
25	210.5	11.2	345	6 Q27951	Q27951 bos taurus
26	210.5	11.2	388	6 Q46523	Q46523 equus zebra
27	210.5	11.2	444	5 Q21966	Q21966 caenorhabdi
28	210	11.2	366	6 Q9N1P5	Q9n1p5 bubalus bub
29	210	11.2	387	11 Q9JUX2	Q9jjx2 rattus norv
30	210	11.2	398	13 P87370	P87370 oncorhynch
31	209	11.1	388	6 Q9N2D3	Q9n2d3 callithrix
32	208	11.1	446	5 Q9N9H3	Q9n9h3 necator ame
33	206	11.0	389	13 Q9W643	Q9w643 gallus gall
34	206	11.0	389	13 Q9PWK1	Q9pwk1 gallus gall
35	204	10.9	413	3 Q14413	Q14413 pichia angu
36	203	10.8	375	6 Q46500	Q46500 bos taurus
37	202.5	10.8	410	5 Q9VQ13	Q9vq13 drosophila
38	201	10.7	632	10 Q9SN13	Q9sn13 arabidopsis
39	200	10.6	383	5 Q76856	Q76856 dictyostell
40	199	10.6	370	6 Q9TTW1	Q9ttw1 bos taurus
41	198	10.5	389	6 Q9MYK3	Q9myk3 sus scrofa
42	197.5	10.5	376	6 Q46499	Q46499 bos taurus
43	197	10.5	389	6 Q9MYK2	Q9myk2 sus scrofa
44	196.5	10.5	450	5 Q76965	Q76965 plasmodium
45	196	10.4	395	5 Q9VEK3	Q9vek3 drosophila

ALIGNMENTS

RESULT 1					
Q9ULS1					
ID Q9ULS1	PRELIMINARY;	PRT;	396 AA.		
AC Q9ULS1;					
DT 01-MAY-2000 (Tremblrel. 13, Created)					
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)					
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)					
DE KIAA1149 PROTEIN (FRAGMENT).					
GN KIAA1149.					
OS Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX NCBI_TaxID=9606;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC TISSUE=BRAIN;					
RA Hiroseawa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;					
RT "Characterization of cDNA clones selected by the GeneMark analysis					
RL from size-fractionated cDNA libraries from human brain.";					
RL DNA Res. 6:329-336(1999).					
DR EMBL; AB032975; BAA8463.1;					
DR HSSP; P56272; IAMS.					
DR INTERPRO; IPR001461;					
DR PFAM; PF000026; asp; 2.					
DR PRINTS; PR00792; PEPSIN.					
FT NON_TER 1					
SQ SEQUENCE 396 AA; 44428 MW; A2CBDCD52DCC089E0 CRC64;					
Query Match	92.9%	Score 1745;	DB 4;	Length 396;	
Best Local Similarity	99.1%	Pred. No. 5.9e-145;			
Matches 330;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;	
Qy 6	YLQAHFTLCGWSSTYRDLRKGVVYPTQGWEGELGTDLVSIHPGPNVTVRANIAATE 65				
Db 1	YLQAHFTLCGWSSTYRDLRKGVVYPTQGWEGELGTDLVSIHPGPNVTVRANIAATE 60				
Qy 66	SDKFFINGSNWEGILGLAYAEIARPDSPFPFSDLVKQTHVNPVLFSLQLCGAGFPLNQ 125				
Db 61	SDKFFINGSNWEGILGLAYAEIARPDSPFPFSDLVKQTHVNPVLFSLQLCGAGFPLNQ 120				
Qy 126	EVLASVGSMTIIGDHSGLYTGSLWYTPIREWYVEIIVRVEINGQDLKMDCKEYNDK 185				

Db 121 EVLASVGGSMIGGIDHSLYTSLWTPYIRREWYEVIIIVRVEINGQDLKMDCKEYNDK 180
QY 186 SIVDSGTTNLRPKKVFEEAAVKSIAASTEKFPDGFNLGEOLVCWQAGTTPWNIFFPVIS 245
Db 181 SIVDSGTTNLRPKKVFEEAAVKSIAASTEKFPDGFNLGEOLVCWQAGTTPWNIFFPVIS 240
QY 246 LYLMEVNTNOSRITTLPOQYLRPEVDVATSDQDCYKFAISQSGTGMVAGVIMEGFYV 305
Db 241 LYLMEVNTNOSRITTLPOQYLRPEVDVATSDQDCYKFAISQSGTGMVAGVIMEGFYV 300
QY 306 FDRARRKRGFAVSACHVHDEFRTAAVEGPFCHL 338
Db 301 FDRARRKRGFAVSACHVHDEFRTAAVEGPFVIL 333
RESULT 2
Q9JL18 PRELIMINARY; PRT; 514 AA.
AC Q9JL18;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE ASPARTYL PROTEASE 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Choi D.K., Sugano S., Sakaki Y.;
RT "Molecular characterization of the mouse Aspl gene, a homolog of the
human Aspl (Down Syndrome Region aspartyl protease).";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF216310; AAF36599.1; -
KW Protease.
SQ SEQUENCE 514 AA; 55799 MW; A70725F2C1DF5B47 CRC64;
Query Match 48.6%; Score 913; DB 11; Length 514;
Best Local Similarity 52.0%; Pred. No. 8.5e-72;
Matches 173; Conservative 55; Mismatches 99; Indels 6; Gaps 3;
QY 3 PFYIQAHTFLCSGWSSTYRDLRKGVVVPYTGKWEGLGTLDLVSIPIHGPVNTVRANIAA 62
Db 118 PHSYIDTYFD--SESSSTYHSGKGFDTVKYTGSGWTGFGVEDLVTPKGFNSFLVNIAT 175
QY 63 ITESDKFFINGNWEGILGLAYAEIARPDSPPEFDSLVKQTHVFNLSQLCGAGFPL 122
Db 176 IFESENFPLGKWNGLGLAYAAALAKPSSSLETFDLSLVAQAKIPDIIFSQMCGAGLPV 235
QY 123 NQSEVLASVGSMTIGIDHSLYTSLWTPYIRREWYEVIIIVRVEINGQDLKMDCKEYN 182
Db 236 AGS---GTNGSGVLVGIEPSLVKGDILWTPYIRREWYEVIIIVRVEINGQDLKMDCKEYN 292
QY 183 YKSDIVDSGTTNLRPKKVFEEAAVKSIAASTEKFPDGFNLGEOLVCWQAGTTPWNIFF 242
Db 293 ADRAIVDSGTTNLRPKKVFEEAAVKSIAASTEKFPDGFNLGEOLVCWQAGTTPWNIFF 352
QY 243 VISLYLMEVNTNOSRITTLPOQYLRPEVDVATSDQDCYKFAISQSGTGMVAGVIMEGF 302
Db 353 KISIVLDRDENASRFRITILPOLYIQPMGAGFNY-ECYRFGISSSTNALVIGATVMEGF 411
QY 303 YVVFDRARRKRGFAVSACHVHDEFRTAAVEGPF 335
Db 412 YVVFDRARRKRGFAVSACHVHDEFRTAAVEGPF 444
RESULT 3
Q9P0D2 PRELIMINARY; PRT; 213 AA.
AC Q9P0D2;
DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE HSPC104 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS cloned from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF161367; AAF28927.1; -
FT NON_TER
SQ SEQUENCE 213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;
Query Match 41.8%; Score 784.5; DB 4; Length 213;
Best Local Similarity 66.0%; Pred. No. 4.4e-61;
Matches 157; Conservative 8; Mismatches 24; Indels 49; Gaps 4;
QY 138 GGIDHSLYTSLWTPYIRREWYEVIIIVRVEINGQDLKMDCKEYNYSIVDSGTTNLR 197
Db 1 GGIDHSLYTSLWTPYIRREWYEVIIIVRVEINGQDLKMDCKEYNYSIVDSGTTNLR 60
QY 198 PKKVFEEAAVKSIAASTEKFPDGFNLGEOLVCWQAGTTPWNIFFVISLYLMEVNTNQS 257
Db 61 PKKVFEEAAVKSIAASTEKFPDGFNLGEOLVCWQAGTTPWNIFFVISLYLMEVNTNQS 120
QY 258 RITILPOQYLRPEVDVATSDQDCYKFAISQSGTGMVAGVIMEGFYVVDRAKRGFAV 317
Db 121 RITILPOQYLRP-----WKMPRPKTTVTVCHLTVIH-----GHY 156
QY 318 SACHVHDEFRTAAVECP-----FCHLGHGRMLQHSR 351
Db 157 GSCY-HGGLPLSLGIPENELACLALACTNSSGRQKWKALCHLGHGRMLQHSR 213
RESULT 4
Q9NZL1 PRELIMINARY; PRT; 396 AA.
AC Q9NZL1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE ASPARTYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Solans A., Estivill X., de la Luna S.;
RT "Cloning of a novel mammalian aspartyl protease.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188277; AAF35836.1; -
KW Protease.
SQ SEQUENCE 396 AA; 43013 MW; 5023A7AF391CEAC9 CRC64;
Query Match 40.9%; Score 767.5; DB 4; Length 396;
Best Local Similarity 52.4%; Pred. No. 3.4e-59;
Matches 150; Conservative 46; Mismatches 75; Indels 15; Gaps 5;
QY 3 PFYIQAHTFLCSGWSSTYRDLRKGVVVPYTGKWEGLGTLDLVSIPIHGPVNTVRANIAA 62
Db 122 PHSYIDTYFD--TERSTYRSGKGFDTVKYTGSGWTGFGVEDLVTPKGFNSFLVNIAT 179
QY 63 ITESDKFFINGNWEGILGLAYAEIARPDSPPEFDSLVKQTHVFNLSQLCGAGFPL 122

Db 180 IFESNFPLPGIKWNGILGLAYATLAKPSSSLETFDLSVLTQANIPNVFSMQMCGAGLPV 239
QY 123 NQSEVLASVGGSMIIIGDHSLYTGSWYTPIRREWYEVILVIRVEINGQDLKMDCKEYN 182
Db 240 AGS---GTNGSLVVGIEFSLYKGDWYTPIKREWYQIEILKEITGGQSLNLDCKEYN 296
QY 183 YKSDIVSGTTLNLRPKKVFEEAAKASSTSEKFPDGFGLGQVLCWQAGTTPWNIFF 242
Db 297 ADKAIVDSGTTLLRLPKQVDFVAVARASLIPEFSDGFMTGWSQACWNTSETPWSYFP 356
QY 243 VISLYLMEVNTQSFRTITLPQQLRVEDVATNSQDCYKF-ALISQ 287
Db 357 KISYLRDENSSRSFRITILPQK-LRVLQ-----CLKPPGLSQ 393

RESULT 5

Q9NZL2
ID Q9NZL2 PRELIMINARY; PRT; 468 AA.
AC Q9NZL2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ASPARTYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Solans A., Estivill X., de la Luna S.;
RT "Cloning of a novel mammalian aspartyl protease";
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188276; AAF35835.1;
KW Protease.
SQ SEQUENCE 468 AA; 50324 MW; 717E0920126A0142 CRC64;

Query Match 38.7%; Score 726; DB 4; Length 468;
Best Local Similarity 43.5%; Pred. No. 1.9e-55;
Matches 145; Conservative 50; Mismatches 82; Indels 56; Gaps 4;

QY 3 PFYIQAHTLCGWSSTYRDLRGVVPYVYQGWEGELGDLVSIPIGPNVTVRANITAA 62
Db 122 PHSYDIYFD--TERSTYSRKSGFDVYKVTQGSMTGFVGEDLVTPKGFNTSELVNIAT 179
QY 63 ITESDKFFINGSNWEGILGLAYABIAIRPDSPPEFFDSLVKQTHVPLNLSQLCGAGFPL 122
Db 180 IFESNFPLPGIKWNGILGLAYATLAKPSSSLETFDLSVLTQANIPNVFSMQMCGAGLPV 239
QY 123 NQSEVLASVGGSMIIIGDHSLYTGSWYTPIRREWYEVILVIRVEINGQDLKMDCKEYN 182
Db 240 AGS---GTNGSLVVGIEFSLYKGDWYTPIKREWYQIEILKEITGGQSLNLDCKEYN 296
QY 183 YKSDIVSGTTLNLRPKKVFEEAAKASSTSEKFPDGFGLGQVLCWQAGTTPWNIFF 242
Db 297 ADKAIVDSGTTLLRLPKQVDFVAVARASLIPEFSDGFMTGWSQACWNTSETPWSYFP 356
QY 243 VISLYLMEVNTQSFRTITLPQQLRVEDVATNSQDCYKF-ALISQ 287
Db 330 -----YIQPMMGAGLNY-ECYRFGISPSSTNALVIGATVMEGF 365
QY 303 YVTFDRARKRIGFAVSACHVHDEFRTAAVEGPF 335
Db 366 YVTFDRARKRIGFAVSACHVHDEFRTAAVEGPF 398

RESULT 6

Q9R1P7
ID Q9R1P7 PRELIMINARY; PRT; 255 AA.
AC Q9R1P7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE ASPARTYL PROTEASE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Accarino M., Fumagalli P., Taramelli R., Ottolenghi S.;
RT "Cloning of a gene from chromosome 21 Down Region encoding a potential transmembrane protease";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF051150; AAD45964.1;
DR INTERPRO; IPR001461;
DR INTERPRO; IPR001969;
DR PFAM; PF00026; asp_2;
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
KW Protease.
FT NON_TER 1
SQ SEQUENCE 255 AA; 28685 MW; 53DE317815996D63 CRC64;

Query Match 29.8%; Score 559.5; DB 11; Length 255;
Best Local Similarity 53.8%; Pred. No. 3.1e-41;
Matches 100; Conservative 34; Mismatches 51; Indels 1; Gaps 1;

QY 150 WYTPIRREWYEVILVIRVEINGQDLKMDCKEYNVDKSIYDSTGNNLRPKKVFEEAAVKS 209
Db 1 WYTPIKREWYQIEILKEITGGQSLNLDCKEYNADKAIVDSGTTLLRLPKQVDFVAV 60
QY 210 KAASSTSEKFPDGFGLGQVLCWQAGTTPWNIFFVISLYLMEVNTNOSFRITLPQYLRP 269
Db 61 ARTSLIPEFSDGFMTGWSQACWNTSETPWFYFKISIVLRDENASRSFRITLPQYLRP 120
QY 270 VEDVATNSQDCYKF-ALISQ 287
Db 121 MNGAGFNY-ECYRFGISPSSTNALVIGATVMEGFYVDFRAQRRVGFVSPCAIEGTFTVS 179
QY 330 AVEGPF 335
Db 180 EISGPF 185

RESULT 7

Q9VRP7
ID Q9VRP7 PRELIMINARY; PRT; 423 AA.
AC Q9VRP7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG6508 PROTEIN.
GN CG6508.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

```

Query Match      13.5%; Score 254; DB 5; Length 423;
Best Local Similarity 28.4%; Pred. No. 4.2e-14;
Matches 93; Conservative 56; Mismatches 132; Indels 46; Gaps 13;

  QY      8 QAHFTLCSSWSTYRDLRKGVVPVTOCKWEGELGTDLVSIHPHGPNVTVRANIAAITESD 67
  DB      113 OKHNKYNSSASSHVDECKGFSIQYSGSLGFLSTDTVDI-DGMVIRNQTFAEADIEPG 171

  QY      68 KEFFKNSWEGILGLAYAEIARPDSPFPFDSLKVQTHVFN-LFSQLQCGAGFPLNQSE 126
  DB      172 SAFVN-TIFDGIIGHAFASISGGVTP---FDNIROGLVKHFVSFYLRDQ--TSQS- 224

  QY      127 VLASVGGSMIGGIDHSLSYTGSLWVTPIRRWYVEVTVIRVEINGODLKMCKEYNVDKS 186
  DB      225 -----GGEVINGGIDRSIYRCINYPVSPMPAYWQFTANSVKIEGILLNCG-----QA 273

  QY      187 IVDSGTNLRPKKVFEAAVSKIA--ASSTEKPPDFGFWLGEOLVCWQAGTTPWTFPVI 244
  DB      274 IADTGTSLIAPLRAYKAINKVLNATDAGDGEAFVDCSSLCR-----LPNV 319

  QY      245 SLVLMGEVNTQSFRTITLPOOYLRPEVEDVATFSODCYK-FAISQSSTGTVMGAVIMEGFV 303
  DB      320 NLNIGTT-----YTLTPKDYIKVQ--ADNNQTLCLSGTYLQGNLLWILGDFLGKYY 372

  QY      304 VVFDRAKRKIGFAVSACHVHDEFRTAA 330
  DB      373 TVFDVGKERIGFA--KLAKKHSSRYVYA 397

RESULT      8
Q9PRG9      Q9PRG9      PRELIMINARY;      PRT;      382 AA.
AC          Q9PRG9;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

```

Query Match 13.5%; Score 253.5; DB 13; Length 382;
Best Local Similarity 23.6%; Pred. No. 4e-14;
Matches 80: Conservative 62; Mismatches 114; Indels 83; Gaps 13;

RESULTS

AC O9VKP6:

GN CG1/134.
OS *Drosophila melanogaster* (Fruit fly).

OC Drosophila melanogaster (Taxa)
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]


```
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003621; AAF52686.1;
DR HSSP; P00797; 2REN.
DR FLYBASE; FBgn0032049; CG13095.
DR INTERPRO; IPR001461;
DR INTERPRO; IPR001969;
DR PFAM; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSTN.
DR PROSITE; PS00141; ASP_PROPEASE;
SQ SEQUENCE 391 AA; 42016 MW; 99A18E7131025E1B CRC64;
```

```
Query Match 13.5%; Score 253.5; DB 5; Length 391;
Best Local Similarity 28.2%; Pred. No. 4.1e-14;
Matches 88; Conservative 54; Mismatches 127; Indels 43; Gaps 12;
```

```
Qy 8 QAHFTLCGNSSTYDLKRGVYVPTQKWECEGLTDLVSTPHGPNVTVRANIAATESD 67
Db 115 QRHNKYDSASSTVYANGEEFAIEYGTGSLGSLNDIVTIA-GISTQNTFGALSEPG 173
Qy 68 KFFINGSNWEIGLGLAYAEIARPDSPFPFDSLKVQTHVPN-LFSLQLCAGGFLNQSE 126
Db 174 TTFVD-APFAGILGLAFSAIA--VDGVTPFPDNNISQGLLDEPVISFYLLKRGQ----- 223
Qy 127 VLASVGSMTIGGDHSLYLTGSLMTPYPIRREWYVEIIVRVEINGDLKMDCKEYNVDKS 186
Db 224 -TAVRGSELIGGLDSSLRYSLTVVPVPSVPAYWQVKVNTIKTNGTLNCGC-----QA 276
Qy 187 IVDSCNNLRPLPKVFEAAVKSIAASSTKFPDGFGLWGEQLV-CWQAGTTPWNIFFVIS 245
Db 277 IADGTSLIAVPLAAYRKINQLGATDND-----GEAFVRCGRVSS-----LPKNV 322
Qy 246 LYLMEVNTNQSFRTITLPPQYLRPVEDVATSDDCYK-FAISQSSTGTVMGAVIMEGVYV 304
```

```
Db 323 LNIIGTV-----FTLAPRDYI--VKVTQNGQTYCMSAFTYMEGLSFWILGDFVIGKFTV 374
Qy 305 VFDRARKRIGFA 316
Db 375 VFDRGNERIGFA 386

RESULT 10
Q9VLK3 PRELIMINARY; PRT; 372 AA.
ID Q9VLK3
AC Q9VLK3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE CGI3095 PROTEIN.
GN CGI3095.
GC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003621; AAF52686.1;
DR HSSP; P00797; 2REN.
DR FLYBASE; FBgn0032049; CG13095.
DR INTERPRO; IPR001461;
DR INTERPRO; IPR001969;
DR PFAM; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSTN.
DR PROSITE; PS00141; ASP_PROPEASE; UNKNOWN 2.
SQ SEQUENCE 372 AA; 40080 MW; D45469E8AD72FCCE CRC64;
```

```
Query Match      12.6%; Score 237.5; DB 5; Length 372;
Best Local Similarity 26.1%; Pred. No. 9.7e-13;
Matches 81; Conservative 44; Mismatches 124; Indels 61; Gaps 9;

QY 15 SGWSSTYRDLRKGVVYPTQGWKGEGLGTDLVSIPIHGNVTVRANAAITESDKFFINGS 74
    |||||
Db 113 SSASTYVANGESFSYQYGTSLTGYLSTDTVDV-NGLSIQSOTFAESINEGTFN-NDA 170
    |||||
QY 75 NWEGILGLAYAEIARPDSPPEFDSLKVQTHVNP-LFSLQLCGAGFPLNQSEVLASVGG 133
    |||||
Db 171 NFDGILGMAYESLA--VDGVAPPYNNMVSGGLVDNSVFSFYLRDQ-----TSTMGG 220
    |||||
QY 134 SMIIGDHSYGLSWYPIREWEYEVIIIVRVEINGDLKMDCKEYNYKSIYDSGTT 193
    |||||
Db 221 ELIFGGSDASLYSCALTYVPISQGYWQFTMAGSSIDGYSLCDDC-----QAIADTGT'S 274
    |||||
QY 194 NLRLP-----KKVFEAAVKSIIKAASSTEKFPDGFGLVQWQAGTTPNIFPVISL 246
    |||||
Db 275 LIVAPYNAVITLSEILNVGEGYLDCCSSVSLPD-----VTFNIGGTNF----- 318
    |||||
QY 247 YLMGEVNTQSFRTILPQQVLRVEDVATSDQDCYKFAISQSSTGTVMGAVIMEGFYVVF 306
    |||||
Db 319 -----VLKPSAVIIQSDGNCMSAFETYMGTDFTWILGDVFIGQYITEF 359
    |||||
QY 307 DRARKRIGFA 316
    |||||
Db 360 DLGNRRIGFA 369

RESULT 11
Q28950 PRELIMINARY; PRT; 380 AA.
ID Q28950;
AC Q28950;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE PREPROCHYMOSIN PRECURSOR (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=GASTRIC;
RA Foltmann B., Jensen A.L., Loenblad P., Smidt E., Axelsen N.H.;
RT "A developmental analysis of the production of chymosin and pepsin in
RT pigs."
RL Comp. Biochem. Physiol. 68B:9-13(1981).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=GASTRIC;
RX MEDLINE=96252892; PubMed=8673731;
RA Houen G., Madsen M.T., Harlow K.W., Loenblad P., Foltmann B.;
RT "The primary structure and enzymic properties of porcine prochymosin
RT and chymosin."
RL Int. J. Biochem. Cell Biol. 28:667-675(1996).
DR EMBL; U14406; AAB08492.1; -.
DR HSSP; P00794; 4CMS.
DR INTERPRO; IPR001461; -.
DR INTERPRO; IPR001969; -.
DR PRAM; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 15 POTENTIAL.
FT PROPEP 16 57 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 58 380 CHYMOSIN.
FT CHAIN 58 380 CHYMOSIN.
SQ SEQUENCE 380 AA; 41771 MW; ECA81B78239D2102 CRC64;
```

```
Query Match      12.5%; Score 235; DB 6; Length 380;
Best Local Similarity 26.0%; Pred. No. 1.7e-12;
Matches 85; Conservative 54; Mismatches 128; Indels 60; Gaps 14;

QY 2 VPFYI-----LQAHFTLCGWSSTYRDLRKGVVYPTQGWKGEGLGTDLVSIPIHGNVTV 56
    |||||
Db 99 VPSVYCKSDACQNHFRNPSKSSSTFNLDKPLSIQYGTSGIGFYDTVMV--AGIVDA 156
    |||||
QY 57 RANIAAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVQTHV--PNLFSLQL 115
    |||||
Db 157 HOTVGLSTQEPSDIITYSEFDGILGLYPELA--SEYTPVPEDNMHRLHVAQDLFAVYM 214
    |||||
QY 116 CGAGFPLNQSEVLASVGGSMIIGDHSYGLSWYPIREWEYEVIIIVRVEINGQDLK 175
    |||||
Db 215 S-----RNDE-----GSMILTLCADPSYITGSLHWVPVMTQLYWQFTVDSVTINGVVA 263
    |||||
QY 176 MD--CKEYNVDSIYDSTTNLRPKVFEAAVKSIIKAASSTEKFPDGFGLVQWQ 232
    |||||
Db 264 CNGGC-----QAILDTGTSLAGPSSDILNTQMAIGATESQ-----YGEFDIDCGS 309
    |||||
QY 233 AGTTNPNIPFVISLYLMGEVNTQSFRTILPQQVLRVEDVATSDQDCY---KFAISQSS 289
    |||||
Db 310 LSSMPTVVF-----EISGRMYPLP--PSAY-----TNDDQGFCTSGFGDGSKS 350
    |||||
QY 290 TGTVMGAVIMEGFYVVFDRARRKIGFA 316
    |||||
Db 351 QHWILGVVFIQEYYSVFDNRANRVGLA 377

RESULT 12
O93428 PRELIMINARY; PRT; 396 AA.
ID O93428;
AC O93428;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE CATHEPSIN D PRECURSOR.
OS Chionodraco hamatus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Notothenioidae; Channichthyidae; Chionodrac.
OX NCBI_TaxID=36188;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Capasso C., Lees W.E., Capasso A., Scudiero R., Carginale V.,
RA Killie P., Kay J., Parisi E.;
RT "Isolation, characterisation and sequencing of a cathepsin D from the
RT liver of the antarctic icefish Chionodraco hamatus."
RL Submitted (JUL-1998) to the EMBL/Genbank/DDJB databases.
DR EMBL; AJ007878; CAA07719.1; -.
DR HSSP; P07339; 1LYB.
DR MEROPS; A01.009; -.
DR INTERPRO; IPR001461; -.
DR INTERPRO; IPR001969; -.
DR PFAM; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
DR Signal; Hydrolyase; Aspartyl protease.
KW SIGNAL 1 18 POTENTIAL.
FT CHAIN 62 396 CATHEPSIN D.
FT CHAIN 62 396 CATHEPSIN D.
SQ SEQUENCE 396 AA; 42662 MW; 62DAF23CCDCF805A CRC64;
```

```
Query Match      12.4%; Score 233; DB 13; Length 396;
Best Local Similarity 28.6%; Pred. No. 2.6e-12;
Matches 91; Conservative 51; Mismatches 124; Indels 52; Gaps 17;

QY 10 HFTLCGWSSTYRDLRKGVVYPTQGWKGEGLGTDLVSIPIHGNVTVRANI--AAITESD 67
    |||||
Db 117 HHKYNKGKSTYVKNGTAFQIAIQYGSLSGLSQDCTI---GDLAIDSLFGEAIKQPG 173
    |||||
```

QY 68 KFFINGSNWEGILGLAYAEIARDPDSPEFFDSLKQTHV-PNLFSLQLCCGAGFPLNOS 126
 Db 174 VAFI-AAKFDGILGMAYPRIS--VDGVAPVFDNIMSKQKVEQNVFS-----FYLNRNP 223
 QY 127 VLASVGSMIIGGDHSLYSGWYPIREWEYEVIIIVRVEINGQDLKM---DCKEYNY 183
 Db 224 D-TEPGGELLGGTDPKYYTGDFNVNVTQAYWIRVDSMAV-GDOLSLCTGCG----- 276
 QY 184 DKSIVDSGTTNLRPKKVFEEAAVKSIAKASSTKEDPDGFWLGEQLVCWQAGTTPWNIFPV 243
 Db 277 -EALVDSGTSLLITGSPSVEVKALQALGA-----FP--LIQGETMV--NCDTVP--SLPV 323
 QY 244 ISLYLMGEVNTQSFRTITLPQOYLPRVEDVATSDQCYKEAI-----SQSSTGTVMGAVI 298
 Db 324 ISFTVGQV-----YTLTGEQILKVTQAGKM--CLSGFMGLDIPAPAGPLWILGDVF 375
 QY 299 MEGFVVVDFDRARRKRGFA 316
 Db 376 MGQYTYVDFDRDANRVGFA 393

RESULT 13

Q9TS27 PRELIMINARY; PRT; 346 AA.
 AC Q9TS27;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
 DE CATHEPSIN D (EC 3.4.23.5).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 ON NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93223670; PubMed=8467789;
 RA Metcalf P., Fusek M.;
 RT "Two crystal structures for cathepsin D: the lysosomal targeting
 RT signal and active site.";
 RL EMBO J. 12:1293-1302(1993).
 DR HSSP; P07339; 1LYB.
 DR INTERPRO; IPR001461; -
 DR INTERPRO; IPR001969; -
 DR PFAM; PF00026; asp. 1.
 DR PRINTS; PR00792; PEPsin.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 SQ SEQUENCE 346 AA; 37711 MW; EBA3376D36D335FE CRC64;

Query Match 12.4%; Score 232; DB 6; Length 346;
 Best Local Similarity 26.8%; Pred. No. 2.6e-12;
 Matches 95; Conservative 63; Mismatches 109; Indels 88; Gaps 19;
 QY 2 VPFIYLO-----AHTLCGWSSTYRDLRKGVY--VPYTGKWEGLTDLVSIPHPG 52
 Db 41 VPSIHKLLDIACWTRHKYNSDSSHY--VKNGTTFDIHYGSGSLYSLQSDTVSPVCPN 98
 QY 53 N-----VTVRANI--AAITSDKFFINGSNWEGILGLAYAEIARDPDSPEFFDSLKQ 104
 Db 99 SSSPGGVTVQRTGFEAIKQPGWFI-AAKFDGILGMAYPRIS--VNNVLPVFDNLMOQ 155
 QY 105 THV-PNLFSLQLCCGAGFPLNQSEVLASVGGSMIIGGDHSLYSGWYPIREWEYEVII 163
 Db 156 KLVDKNVFS-----FFLNR-DPKAOPGGELMGGTDSKYRGSLSLHFHNVTRQAYWQIH 207
 QY 164 IVRVEINGQDLKMDCKEYNDKSIIVDSGTTNLRPKKVFEEAAVKSIAKASSTKEDPDGFW 223
 Db 208 MDQUDV-GSSLTV-CK--GCCAIVDGTSLVGPVEVRELQKAVPLIQ----- 256
 QY 224 LGEQLV-CWQAGTTPWNIFVSIYLMGEVNTQSFRTITLPQOYLPRVEDVATSDQCYK 282
 Db 257 -GEWIPCEKVS-----LPEVIVKLG-----KDYALSPED-YA 289

QY 283 FAISQSSTGT-----VMGAVIMEGFVVVDFDRARRKRGFAVSA 319
 Db 290 LKVSQAEITVCLSGFMDMDIPPGCPWLWILGDVFIGRYTYVDFDRDQNRVGLAEAA 344
 RESULT 14
 Q91322 PRELIMINARY; PRT; 384 AA.
 AC Q91322;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
 DE PEPsinOGEN PRECURSOR.
 GN PEPsinOGEN.
 OS Rana catesbeiana (Bull frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranidae; Rana.
 ON NCBI_TaxID=8400;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92042186; PubMed=1939266;
 RA Yakabe E., Tanji M., Ichinose M., Goto S., Miki K., Kurokawa K.,
 RT Ito H., Kageyama T., Takahashi K.;
 RT "Purification, characterization, and amino acid sequences of
 RT pepsinogens and pepsins from the esophageal mucosa of bullfrog (Rana
 RT catesbeiana).";
 RL J. Biol. Chem. 266:22436-22443(1991).
 DR EMBL; M73750; AAA49530.1; -
 DR HSSP; P20142; IAVF.
 DR MEROPS; A01.003; -
 DR INTERPRO; IPR001461; -
 DR INTERPRO; IPR001969; -
 DR PFAM; PF00026; asp. 1.
 DR PRINTS; PR00792; PEPsin.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Signal; Hydrolase; Aspartyl protease.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 53 384 PEPsin.
 SQ SEQUENCE 384 AA; 41764 MW; D3441A80EF6F0BB1 CRC64;

Query Match 12.4%; Score 232; DB 13; Length 384;
 Best Local Similarity 24.8%; Pred. No. 3.1e-12;
 Matches 86; Conservative 53; Mismatches 124; Indels 84; Gaps 14;
 QY 2 VPFIYLOA-----HFTLCGWSSTYRDLRKGVYVYTGKWEGLTDLVSIPHPNVTV 56
 Db 93 VPSTYCQSQACTNHPQFNPSQSSSYSSNQOQFSLQYGTGSLTGILGYDTVQI----- 144
 QY 57 RANIA-----AITSDKFFINGSNWEGILGLAYAEIARDPDSPEFFDSLKQTHVP 108
 Db 145 -ONTAISQOEFGLSVTEPGTNFVY-AQFDGILGLAYPSIA--EGGATVMOGMQINLIN 200
 QY 109 N-LFSLQLCCGAGFPLNQSEVLASVGGSMIIGGDHSLYSGWYPIREWEYEVIIVRV 167
 Db 201 QPLFAFYLSQOONSQ-----GGEVAFGVQDQNTYSGQIYWIPTVSETYWGIGQGF 252
 QY 168 EINGQD---LKMCKEYNDKSIIVDSGTTNLRPKKVFEEAAVKSIAKASSTKEDPDGFWL 224
 Db 253 SVNGQATGWCSSQGC-----QGIVDTGTSLLTAPQSVFSSLMQSIGAQDQN----- 298
 QY 225 GEQLV-CWQAGTTPWNIFPVI-----SLYLMGEVNTQ---SFRITILPQOYLPRVED 272
 Db 299 GOYAVCSNIOQLSTISFTISGVSFPLPPSAYVLQONSQGYCTIGTIPTYLPSQOQL-- 356
 QY 273 VATSQDDCYKFAISQSTGTVMGAVIMEGFVVVDFDRARRKRGFAVSA 319
 Db 357 -----WILGDVFLRQYYSYVDYLDGNNQVGAFAA 384
 RESULT 15
 Q29080

ID Q29080 PRELIMINARY; PRT: 385 AA.
AC Q29080:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE PEPsinogen A PRECURSOR.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89174702; PubMed=2494172;
RA Lin X.L., Wong R.N.S., Tang J.;
RT "Synthesis, purification,, and active site mutagenesis of recombinant
RT porcine pepsinogen.,";
RL J. Biol. Chem. 264:4482-4489(1989).
DR EMBL; J04601; AAA31096.1; -.
DR HSSP; P00791; IPSA.
DR INTERPRO; IPR001461; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001969; -.
DR PFAM; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Signal; Hydrolase; Aspartyl protease.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 385 POTENTIAL.
FT CHAIN 60 385 POTENTIAL.
SQ SEQUENCE 385 AA; 41310 MW; 9ABC406AD5877708 CRC64;

Query Match 12.3%; Score 230.5; DB 6; Length 385;
Best Local Similarity 25.8%; Pred. No. 4.2e-12;
Matches 85; Conservative 60; Mismatches 125; Indels 59; Gaps 15;

QY 2 VPTIYLAHFTLCGWN-----SSTYRLRGVYVYPTQGWEGELGTLVSIHPGNV 54
DB 99 VPSVYCSS--LACSDHNFNPDDSTFEATSQELSTITYGTGSMGTGLGYDTYQV---GGI 153
QY 55 TVRANTAAITESDK-FFINGSNNEGILGLAYAEIARPDSPPEPFDLSLVKQHV-PNLFS 112
DB 154 SDTNQIFGLSETEPGSLYAPFDGILGLAYPSISA--SCATPVFDNLWDQGLVSQDLFS 211
QY 113 LOLCGAGFPLNQSEVLASVGSMLIGDHSYLTGSLWYTPIRREWYEVIIIVRVEINGQ 172
DB 212 VYLSS-----NDDS-----GSVVLGGIDSSYTGSLNWPVSVGEYQWITLDSITMDGE 261
QY 173 DLKMD--CKEYNDKSIDVSGTTLNRLPKVFEAAVKSIAASSTKFKPDPGFWLGEOLV- 229
DB 262 TIACSGGC-----QAIVDGTSLTGTPTSAIANIQSDIGASENS-----YGEWVIS 307
QY 230 CWQAGITPNWIFPVISLYLMGEVNTNSFRITILPQOYLRPVEDVATSQDDCYKFAISQSS 289
DB 308 CSSIDSLPDIIVF-----TINGVQPLSPSAYILQDDDSCTSG---FEGMDVPTS 353
QY 290 TGT--VMGAVIMEGFYVVFDRARKRIGFA 316
DB 354 SGEILWLGDFVFIQYTYVDFRANKVGLA 382

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:21:30 ; Search time 44.26 Seconds
(without alignments)
538.480 Million cell updates/sec

Title: US-09-215-435-167

Perfect score: 1878

Sequence: 1 MVPFTYLOAHFTCLSGWST.....EGPFCHLGHRLQLHSTDR 351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 6790655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_66:*

1: pirl:*

2: pirl2:*

3: pirl3:*

4: pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1673	89.1	501	2 A59090	aspartic proteinas
2	253.5	13.5	367	1 PECH	pepsin A (EC 3.4.2
3	253.5	13.5	382	2 JE0370	pepsin A (EC 3.4.2
4	251	13.4	388	1 S19682	pepsin A (EC 3.4.2
5	239	12.7	387	2 B38302	pepsin (EC 3.4.23.
6	237.5	12.6	387	2 E38302	pepsin (EC 3.4.23.
7	235	12.5	380	2 I47176	chymosin (EC 3.4.2
8	234.5	12.5	388	1 S19684	pepsin A (EC 3.4.2
9	234	12.5	386	1 PEPG	pepsin A (EC 3.4.2
10	233	12.4	398	2 S66465	cathepsin E (EC 3.
11	232	12.4	384	2 A39314	gastricsin (EC 3.4
12	231.5	12.3	388	1 PEHU	pepsin A (EC 3.4.2
13	231.5	12.3	388	1 PEMQAR	pepsin A (EC 3.4.2
14	230.5	12.3	383	2 A41443	pepsin (EC 3.4.23.
15	230	12.2	387	2 C38302	pepsin (EC 3.4.23.
16	230	12.2	396	2 A34401	cathepsin E (EC 3.
17	229.5	12.2	388	1 PEMQAJ	pepsin A (EC 3.4.2
18	229.5	12.2	391	2 A43356	cathepsin E (EC 3.
19	229	12.2	387	2 D38302	pepsin (EC 3.4.23.
20	227.5	12.1	388	2 A30142	pepsin A (EC 3.4.2
21	227	12.1	381	1 CMSHB	chymosin (EC 3.4.2
22	225.5	12.0	387	2 JC7245	pepsinogen A - com
23	225.5	12.0	388	2 B30142	pepsin A (EC 3.4.2
24	225	12.0	381	2 JC7247	prochymosin - comm
25	224	11.9	412	1 KHHUD	cathepsin D (EC 3.
26	220	11.7	396	2 S36865	cathepsin E (EC 3.
27	219	11.7	381	1 CMBO	chymosin (EC 3.4.2
28	218	11.6	377	1 PEMOCJ	gastricsin (EC 3.4
29	218	11.6	389	2 A38302	pepsin (EC 3.4.23.

ALIGNMENTS

RESULT 1

A59090

aspartic proteinase (EC 3.4.23.-) BACE precursor - human

N;Alternate names: beta-secretase; beta-site APP cleaving enzyme

C;Species: Homo sapiens (man)

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 11-May-2000

C;Accession: A59090

R;Vassar, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.; Tepl

M.A.; Biere, A.L.; Curran, E.; Burgess, T.; Louis, J.C.; Collins, F.; Treanor, J.; Ro

Science 286, 735-741, 1999

A;Title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran

A;Reference number: A59090; MUID:20002972

A;Note: submitted to Genbank, September 1999

A;Accession: A59090

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-501 <VAS>

A;Cross-references: GB:AF190725; NID:g6118538; PIDN:AAF04142.1; PID:g6118539

C;Genetics:

A;Gene: BACE

C;Superfamily: beta-secretase

C;Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-45/Domain: propeptide #status predicted <PRO>

F;46-501/Product: acid proteinase BACE #status predicted <MAT>

F;461-477/Domain: transmembrane #status predicted <TRN>

F;93,289/Active site: Asp #status predicted

F;153,172,223,354/Binding site: carboxylate (Asn) (covalent) #status predicted

F;330-380/Disulfide bonds: #status predicted

Query Match 89.1%; Score 1673; DB 2; Length 501;

Best Local Similarity 95.3%; Pred. No. 4.2e-137;

Matches 322; Conservative 1; Mismatches 7; Indels 8; Gaps 2;

QY 3 PFI--YLOAHFTCLSGWSSYRDLRKGYVYPTQKGEGELGTDLVSPHGNVTVRANI 60

Db 107 PFLRYQRQQL-----SSTYRDLRKGYVYPTQKGEGELGTDLVSPHGNVTVRANI 160

QY 61 AAITESDKFFINGNWEGILGLAYAEIARPDSPPEPFDLSLVKQTHVFNLSLQCGAGF 120

Db 161 AAITESDKFFINGNWEGILGLAYAEIARPDSPPEPFDLSLVKQTHVFNLSLQCGAGF 220

QY 121 PLNOSSEVLASVGGSMIIIGGDHSLYTGSLWYTPIRREWYVEIIVRVVEINGDLKMDCKE 180

Db 221 PLNOSSEVLASVGGSMIIIGGDHSLYTGSLWYTPIRREWYVEIIVRVVEINGDLKMDCKE 280

QY 181 YNYDKSVDSGTTNLRPLPKKVFEEAAVKSIIKAASSTKFPDGFVWLGQVLCVQAGTTPWNI 240

Db 281 YNYDKSVDSGTTNLRPLPKKVFEEAAVKSIIKAASSTKFPDGFVWLGQVLCVQAGTTPWNI 340

QY 241 FPVLSILYMGVETNQSPFRITILPQOYLPRPVEDVATSDQDCYKFAISQSSTCTVVGAVIME 300

[illegible]

Db 123 SSTYRATSKTVSYTGSGMTGILGYDVKV---GGISDNTQIQLSETPGFFLYPAPF 179
QY 77 EGILGLAYAEIARDDSPPEFFDLSLVKQTHV-PNLSLQLCGAGFPPLNQSEVLASVGGSM 135
Db 180 DGLGLAYPSIS--SSGATPVFNINQRLVSODLFSVLSAD-----DQS-----GSVV 227
QY 136 IIGGIDHSLTSGSLWYTPIRREWYEVIIIVRVEINGDL--KMDCKEYNDKSIIVDSGTT 193
Db 228 IFGGIDSSYTGSLWNVPVSEGVQWISVDSITMNGKTIACAKG-----QAIVDTGTS 281
QY 194 NLRLPKKVFEEAVKSIKAASSTKEFPDGLWGEOLV-CWQAGTTPWNIPFVLSLYLMGEV 252
Db 282 LLTGPTSPIANIQSDIGASENSD-----GEMVWSCSAISLDPDVF----- 322
QY 253 TNSQFRITILPOQY-LRPVEDVATSDQCYK-----FAISQSSTGTVMGAVIMEGFVVVF 306
Db 323 -----TINGVQYPLPPSAYILQSOGSCTSGFGQMDVPTESGELWILGDVFIQRYPTVF 375
QY 307 DRARKRIGFA 316
Db 376 DRANNQVGLA 385

RESULT 5
B38302
pepsin (EC 3.4.23.-) II-1 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
R:Kageyama, T.; Tanabe, K.; Koizumi, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucle
A:Reference number: A38302; MUID:91009127
A:Accession: B38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.7%; Score 239; DB 2; Length 387;
Best Local Similarity 26.3%; Pred. No. 5.3e-13;
Matches 83; Conservative 56; Mismatches 114; Indels 62; Gaps 14;

QY 18 SSTYRDLRKGVYVPTQGWEGELGTLVSIPIHGPNTVVRANTAAITSD---KFFINGS 74
Db 122 SSTFQATSETLSITYGTGSGMTGILGYDVKV---GNIEDTQIIFGLSKTEPGITFLV--A 176
QY 75 NWEGILGLAYAEIARDDSPPEFFDLSLVKQTHV-PNLSLQLCGAGFPPLNQSEVLASVGG 133
Db 177 PFDGILGLAYPSISASAT--PVFDNMWNEGLVSEDLFSVYLSNG-----EKGS 224
QY 134 SMITGGIDHSLTSGSLWYTPIRREWYEVIIIVRVEINGDLK--DCKEYNDKSIIVDSG 191
Db 225 MVMEFGIDSSYTGSLWNVPVSEGVQWISVDSITMNGKTIACADSC-----QAVVDTG 278
QY 192 TTNLRLEPKKVFEEAVKSIKAASSTKEFPDGLWGEOLV-CWQAGTTPWNIPFVLSLYLMG 250
Db 279 TSLLAGTSAISKSIQYIGASKNL-----LGENIISCAISDLPDVF----- 321
QY 251 EVTNQSFRTILPOQYLRPVED-VATSDQDC---YKFAISQSSTGT--VMGAVIMEGFVV 304
Db 322 -----TINNVOYPLPASAYILKEDDDCLSGFGDMNLDTSYGELWILGDVFIQRYFT 372
QY 305 VFDRARKRIGFAVSA 319
Db 373 VFDRANNQVGLAAAA 387

RESULT 6
E38302

pepsin (EC 3.4.23.-) IIis precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 22-Jun-1999
C:Accession: E38302
R:Kageyama, T.; Tanabe, K.; Koizumi, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: E38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M5237; GB:J05639; NID:g165597; PIDN:AAA85370.1; PID:g165598
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.6%; Score 237.5; DB 2; Length 387;
Best Local Similarity 26.0%; Pred. No. 7.2e-13;
Matches 87; Conservative 62; Mismatches 121; Indels 65; Gaps 15;

QY 2 VPFIYLAHFTLCGW-----SSTYRDLRKGVYVPTQGWEGELGTLVSIPIHGPV 54
Db 101 VPSVYCSS--AACSVMHQFNPEDSSTFQATSESLITYGTGSMTGFLGYDVKV---GNI 155
QY 55 TVRANIAAITESDK-EFINGSNWEGILGLAYAEIARDDSPPEFFDLSLVKQTHV-PNLS 112
Db 156 EDNQIFGLSESEPGSFLYAPFDGILGLAYPSISSDAT--PVFDNMWNEGLVSEDLFS 213
QY 113 LQLCGAGFPPLNQSEVLASVGGSMIIGGIDHSLTSGSLWYTPIRREWYEVIIIVRVEINGQ 172
Db 214 VYLSDD-----ESGVVWFGGIDSSYTGSLWNVPVSEGVQWISVDSITMDGE 263
QY 173 DLKM--DCKEYNDKSIIVDSGTTNLRPKKVFEEAVKSIKAASSTKEFPDGLWGEOLV 230
Db 264 TIACADSC-----QAIVDTGTSLLAGP---TSAISNIQSYIGASENSD---GEMIVS 309
QY 231 WQAGTTPWNIPFVLSLYLMGEVNTQSFRTILPOQYLRPVEDVATSDQDC---YKFAIS 286
Db 310 CS-----SMISLPNIV-----FTINGVQYVPASAYILEDDACISGFEGMNL 352
QY 287 QSSTGT--VMGAVIMEGFVVFDRAKRIGFAVSA 319
Db 353 DTYTGELWILGDVFIQRYFTVFDRAANNQGLAAAA 387

RESULT 7
I47176
chymosin (EC 3.4.23.4) precursor - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 20-Jun-2000
C:Accession: I47176
R:Foltmann, B.; Jensen, A.L.; Loenblad, P.; Smidt, E.; Axelsen, N.H.
Comp. Biochem. Physiol. A 68, 9-13, 1981
A:Title: A Developmental Analysis of the Production of Chymosin and Pepsin in Pigs.
A:Reference number: I47176
A:Accession: I47176
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-380 <FOL>
A:Cross-references: EMBL:U14406; NID:g540096; PIDN:AAB08492.1; PID:g540097
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase

Query Match 12.5%; Score 235; DB 2; Length 380;
Best Local Similarity 26.0%; Pred. No. 1.2e-12;
Matches 85; Conservative 54; Mismatches 128; Indels 60; Gaps 14;

QY 2 VPFIY-----LQAHFTLCGWSSTYRDLRKGVYVPTQGWEGELGTLVSIPIHGPVTV 56
Db 99 VPSVYCKSDACQNHHRFNPSKSTTFQNLKPLSIQYGTGSIQGLGYDVTVMV--AGIVDA 156

QY 234 GTTPWNIFPVISLYLMGEVYNQSFRTITLPPQY-LRPVEDNATSQDDCYR-----FATSQ 235

QY 234 GTTPWNIFPVISLYLMGEVYNQSFRTITLPPQY-LRPVEDNATSQDDCYR-----FATSQ 235


```
RESULT 2
US-09-434-427-2
; Sequence 2, Application US/09434427
; Patent No. 6162630
; GENERAL INFORMATION:
; APPLICANT: POWELL, DAVID J.
; APPLICANT: SOUTHAN, CHRISTOPHER
; APPLICANT: CHAPMAN, CONRAD G.
; APPLICANT: EVANS, JOANNE R.
; TITLE OF INVENTION: ASP1
; FILE REFERENCE: GH-70262-D1
; CURRENT APPLICATION NUMBER: US/09/434,427
; CURRENT FILING DATE: 1999-11-04
; EARLIER APPLICATION NUMBER: US 08/999,723
; EARLIER FILING DATE: 1997-10-06
; EARLIER APPLICATION NUMBER: UK 9626022.9
; EARLIER FILING DATE: 1996-12-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-434-427-2

Query Match 49.3%; Score 925; DB 3; Length 518;
Best Local Similarity 52.3%; Pred. No. 9.9e-95;
Matches 174; Conservative 56; Mismatches 97; Indels 6; Gaps 3;

QY 3 PRIYLAHFTLCGWSSTYRDLKRGVYVPTQGWEGELGTLVSIPIHGNVTVRANIAA 62
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 122 PHYSDTYFD--TERSTYRSGEDVTKYTOGWTGFGVDELVTIPKGFNTSLVNIAT 179
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 63 ITESDKFFNGSNWEGILGLAYAEIARDDSPDPFDSLVKQTHVNPVNFSLQCGAGPL 122
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 180 IFESENFLPGIKWNGILGLAYATLAKPSSLETFFDSLVTOANIPNVFQMCGAGLPV 239
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 123 NOSEVLASVGGSMITGGIDHSLYTGLSWTPTRRREWYEVIIIVRVINGODLKMCKEYN 182
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 240 AGS---GTNGGSLVGGIEPSLYKGDITWTPKREWYQIIEILKLEIGQSUNLDCREYN 296
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 183 YDKSVSDGTTNLRPKPKVFEAAVKSIAKASTKFPDGFGLGQVLCVQAGTTTPWNIFP 242
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 297 ADKAVSDGTTLLRLPKQKVDAAVEAVARASLIPEFSDGFWTGSOLACWTNSETPWSYFP 356
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 243 VLSLYLMGEVNTQSFRTITLPOQYLRPVEDVATSDQDCYKFAISQSSGTCTVMGAVIMEGF 302
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 357 KISIYLRDENSRSFRITILPOLYTOPMAGLNY-ECYRFGISPTNALVIGATVMEGF 415
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 303 YVVFDRARKRIGFAVSACHVHDEFTAAVEGPF 335
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 416 YVIFDRAQRVGFAGFAASPCAEIAGAAVSEISGPF 448
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```
RESULT 3
US-08-208-007A-13
; Sequence 13, Application US/08208007A
; Patent No. 5501969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM-ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
```

```
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,007A
; FILING DATE: March 8, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: No. 5501969e
; FILING DATE: No. 5501969e
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-208-007A-13

Query Match 12.2%; Score 230; DB 1; Length 396;
Best Local Similarity 25.2%; Pred. No. 3.2e-17;
Matches 84; Conservative 53; Mismatches 134; Indels 62; Gaps 14;

QY 2 VPFYI-----LQAHTLCGWSSTYRDLKRGVYVPTQGWEGELGTLVSIPIHGNVTV 56
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 104 VPSVYCTSPACKTHSRFPQSOSSTYVQPSFSIQYGTGSLGIIAGDQVSV-EGLTVVG 162
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 57 RANIAATSDKFFNGSNWEGILGLAYAEIARDDSPDPFDSLVKQTHVNPVNFSLQIC 116
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 163 QQFGESVTEPGQTFVD-AEFDGILGLGYPSLA--VGGVTPVFDNMQA-----NLVDLPMF 215
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 117 GAGPFLNGSEVLASVGGSMITGGIDHSLYTGLSWTPTRRREWYEVIIIVRVINGODLKM 176
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 216 SVYMSNPE---GGAGSELIFGGYDHSFSLWVPTKQAYWQIALDNIQVGG--TVM 270
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 177 DCKEYNDKSVSDGTTNLRPKPKVFEAAVKSIAKASTKFPDGFGLGQVLCVQAGTT 236
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 271 FCSE--GCAIVDTGCTSLITGTPSDKIKQLQNAIGAAP-----VDGEYAVE-----CA 315
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 237 PWNTPPVLSLYLMGEVNTQSFRTITLPOQYLRPVEDVATSDQDCYKFAISQSSGTG----- 291
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 316 NLNVMPDVTFTING-----VPYTLSPATAY--TLDDFVDMQWQFC-----SSGFQGLD 359
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 292 -----TVMGAVIMEGFYVVFDRARKRIGFA 316
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 360 IHPPAGPLWILGDVFIQFYVDFDRGNRVGLA 392
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 4
5217891-15
; Patent No. 5217891
; APPLICANT: BRAKE, ANTHONY J.;VAN DEN BERG, JOHAN A.
; TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYVEROMYCES
; A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS
; POLYPEPTIDES
; NUMBER OF SEQUENCES: 23
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/507,398
; FILING DATE: 09-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 78,551
; FILING DATE: 28-JUL-1987
; SEQ ID NO:15:
; LENGTH: 458
5217891-15
```

Query Match	12.1%	Score 227;	DB 5;	Length 458;
Best Local Similarity	26.0%;	Pred. No. 8.7e-17;		
Matches	85;	Conservative 59;	Mismatches 123;	Indels 60; Gaps 14;
QY	2	VPIYIQA-----HFTLCGWSSTYRDLRGVYVYPTQGWKEGLGFDLSVIPHGPNVT	56	
		:	:	:
Db	177	VPSYCKSNACKNRHQRDFDKSKSTFQNLGRDLSTHYGTGSMQGLIGYDTVTVSINVIDIQ	236	
QY	57	RANIAITSDKFFINGSNWEGLILGLAYAIARPDDSPPEFFDSLVKQTHV-PNLFSLQL	115	
		: : :	: : :	: : :
Db	237	TVGLSTGEPGDVF--TYAEPDGLIGMAYPSLA--SEYSIPVFNMMNRHLVAQDLFSYIM	292	
QY	116	CGAGFPLNGSEVLASVGGSMIGGIDHSLTGSILWYTPIRRENYEYIIVRVEINGODLK	175	
		: :	: :	: :
Db	293	DRNG-----QESMT-----LGAIDPSYTGSLUHWVPVTVQYQWQTVDSVTISGVVVA	341	
QY	176	MD--KREYNKYDSDTGNLRPKKVFEEAAKSIKAASSTEFKPDGFWLGE-QLVCWQ	232	
		: : : :	: : : :	: : : :
Db	342	CEGCG-----QALDTGTSKLVCPSSDIILNIQAIGATQK-----YGEFDIDCDN	387	
QY	233	AGTTPNNIFPISLYLMGEVNTNSFRITLPQOYLRPEVDVATSQDCY---KFAISQSS	289	
		: : :	: : :	: : :
Db	388	LSYMPVWF-----EINGKMYPLT--PSAY-----TSQDQGFCTSGFQSENHS	428	
QY	290	TGTVMGAVIMEGFYVVFDFDRKKRIGFA	316	
		: : :	: : :	: : :
Db	429	QGWILGDVFIREYYSVDFRANNLVGLA	455	

RESULT 6
 US-08-846-021A-8
 ; Sequence 8, Application US/08846021A
 ; Patent No. 5948682
 ; GENERAL INFORMATION:
 ; APPLICANT: Moloney, Maurice M.
 ; TITLE OF INVENTION: Preparation of Heterologous Proteins on
 ; TITLE OF INVENTION: Oil Bodies
 ; NUMBER OF SEQUENCES: 31
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BERESKIN & PARR
 ; STREET: 40 King Street West
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5H 3Y2
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA: US/08/846.021A
 ; APPLICATION NUMBER: US/08/846.021A
 ; FILING DATE: April 25, 1997
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gravelle, Micheline
 ; REGISTRATION NUMBER: 40,261
 ; REFERENCE/DOCKET NUMBER: 9369-039
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 364-7311
 ; TELEFAX: (416) 361-1398
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 427 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-846-021A-8

```

RESULT 7
US-08-088-633-2
; Sequence 2, Application US/08088633
; Patent No. 5324660
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: Genes which Influence Pichia Proteolytic
; TITLE OF INVENTION: Activity, and Uses therefor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 South LaSalle Street, Suite 900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/088,633
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/678,916
; FILING DATE: 01-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E
; REGISTRATION NUMBER: 31192
; REFERENCE/DOCKET NUMBER: 50848
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)552-1311
; TELEFAX: (619)552-0095
; TELEX: 20 6566 PATLAW CGO
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-088-633-2

```

RESULT 8
 US-08-245-756-2
 ; Sequence 2, Application US/08245756
 ; Patent No. 5541112
 ; GENERAL INFORMATION:
 ; APPLICANT: Gleeson, Martin A
 ; APPLICANT: Howard, Bradley D
 ; TITLE OF INVENTION: GENES WHICH INFLUENCE PICICHA PROTEOLYTIC
 ; TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
 ; STREET: 444 So. Flower Street, Suite 2000
 ; CITY: Los Angeles
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/245,756
 ; FILING DATE: 16-May-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/088,633
 ; FILING DATE: 06-JULY-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/678,916
 ; FILING DATE: 01-APR-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Seidman, Stephanie
 ; REGISTRATION NUMBER: 33,779
 ; REFERENCE/DOCKET NUMBER: 9763
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619-546-4737
 ; TELEFAX: 619-546-9392
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 410 amino acids


```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-245-756-2

Query Match 10.0%; Score 188; DB 1; Length 410;
Best Local Similarity 24.1%; Pred. No. 1.7e-12;
Matches 74; Conservative 40; Mismatches 147; Indels 46; Gaps 9;

QY 18 SSTYRDLKGVVYPYTGKWEGLGTLVSIHPGPNVTVRANIAAITESSDKFFINGSNWE 77
Db 143 SSTYKNGSSFEIRYSGSGMEGVSDVLOI--GDLTPKVDFAEATSEPLAFAGKFD 200

QY 78 GILGLAYAEIARPDSPPEPFDLSLVKQTHVPNLF---SLQLCGA---GFLPNQSEVLASV 131
Db 201 GILGLAY-----DSISVKNIVPIYKALELDDLDEPKFAFYLGDTDKDESD 246

QY 132 GSGMIIGDHSYTGSLWTPTRREWEYEVIIYRVINGODLKMCKEYNDKSIYVDSG 191
Db 247 GGLATGGVDKSYEGKITWLPVRRKAYE-----VSFGVGLGSEYAELOKGTGAADTG 301

QY 192 TTNLRPKPKVFEAAVKSIAASSTKFPDGFGLGEQLVCWQAGTTPWNIPFVISLYLMGE 251
Db 302 TSLIALPSGLAEILNAEIGATG-----WSQYAVDCDTRDS---LPDLTLTFAG- 348

QY 252 VTNQSFRTITLPOQYLRPVEDVATSDDCYKFAISQSTGTVMGAVIMEGFYVVFDRARK 311
Db 349 -----YNFTITPYDYTLLEVSGSCISAFTPMDFP-EPIGPLAIIGDSFLRKYSYVDLGKD 402

QY 312 RIGFAVS 318
Db 403 AVGLAKS 409

RESULT 9
US-08-441-750-2
; Sequence 2, Application US/08441750
; Patent No. 5691166
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,750
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,756
; FILING DATE: 16-May-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/088,633
; FILING DATE: 06-JULY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/678,916
; FILING DATE: 01-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
```

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/088,633
;; FILING DATE: 06-JULY-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/678,916
;; FILING DATE: 01-APR-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 9763
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-238-0999
;; TELEFAX: 619-238-0062
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 410 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-441-751-2

Query Match 10.0%; Score 188; DB 2; Length 410;
Best Local Similarity 24.1%; Pred. No. 1.7e-12;
Matches 74; Conservative 40; Mismatches 147; Indels 46; Gaps 9;
QY 18 SSTYRDLKGVVPTQGWEGELGTLVSIPIHGPNTVVRANIAAITESDKFFINGSNWE 77
Db 143 SSTYKNGSSFEIRVSGSMGVSDVQLI--GDLTPKVDFAETSEPGLAFAFGFD 200
QY 78 GILGLAYAEIARPDSPPEFFDSLKVQTHVPNLF---SLQLCGA---GFPLNQSEVLASV 131
Db 201 GILGLAY-----DSISVKNIVPPIYKALELDLDEPKFAFYLGDTOKDESD 246
QY 132 GSGMIIGDHSILGSLWYTPIRREWYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSG 191
Db 247 GGLATFGGVGDKSKYEGKITWLPVRRKAYWE-----VSFDGVLGSEYAELOKTAIDTG 301
QY 192 TTNLRPKKVFEEAAVKSIAASSTEFKPDGFWLGEOLVCWQAGTTPWNIFPVISLYLME 251
Db 302 TSLIALPSGLAEILNAEIGATKG-----WSGQYAVDCDTRDS----LPDLTLTFAG- 348
QY 252 VTNQSFRTILPQQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYVVFDRARK 311
Db 349 -----YNTITPYDVTLEVSGSCISAFPMDFP-EPIGPLAIGDSFLRKYSVYDLGKD 402
QY 312 RIGFAVS 318
Db 403 AVGLAKS 409

RESULT 11
PCT-US92-02521-2
; Sequence 2, Application PC/TUS9202521
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 South LaSalle Street, Suite 900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: PCT/US92/02521
;; FILING DATE: 19920321
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/678,916
;; FILING DATE: 01-APR-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 50848PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619)552-1311
;; TELEFAX: (619)552-0095
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 410 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US92-02521-2

Query Match 10.0%; Score 188; DB 4; Length 410;
Best Local Similarity 24.1%; Pred. No. 1.7e-12;
Matches 74; Conservative 40; Mismatches 147; Indels 46; Gaps 9;
QY 18 SSTYRDLKGVVPTQGWEGELGTLVSIPIHGPNTVVRANIAAITESDKFFINGSNWE 77
Db 143 SSTYKNGSSFEIRVSGSMGVSDVQLI--GDLTPKVDFAETSEPGLAFAFGFD 200
QY 78 GILGLAYAEIARPDSPPEFFDSLKVQTHVPNLF---SLQLCGA---GFPLNQSEVLASV 131
Db 201 GILGLAY-----DSISVKNIVPPIYKALELDLDEPKFAFYLGDTOKDESD 246
QY 132 GSGMIIGDHSILGSLWYTPIRREWYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSG 191
Db 247 GGLATFGGVGDKSKYEGKITWLPVRRKAYWE-----VSFDGVLGSEYAELOKTAIDTG 301
QY 192 TTNLRPKKVFEEAAVKSIAASSTEFKPDGFWLGEOLVCWQAGTTPWNIFPVISLYLME 251
Db 302 TSLIALPSGLAEILNAEIGATKG-----WSGQYAVDCDTRDS----LPDLTLTFAG- 348
QY 252 VTNQSFRTILPQQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYVVFDRARK 311
Db 349 -----YNTITPYDVTLEVSGSCISAFPMDFP-EPIGPLAIGDSFLRKYSVYDLGKD 402
QY 312 RIGFAVS 318
Db 403 AVGLAKS 409

RESULT 12
US-08-360-673-6
; Sequence 6, Application US/08360673
; Patent No. 5679544
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Fournier, Alain
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLOVYEROMYCES YEASTS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

Best Match

RESULT 13
US-08-115-753-1
; Sequence 1, Application US/08115753
; Patent No. 6017762
; GENERAL INFORMATION:
; APPLICANT: JARA, Patrick
; APPLICANT: LEGOUX, Richard
; APPLICANT: LOISON, Gerard
; APPLICANT: RAZANAMPARANY, Voahangy
; TITLE OF INVENTION: Cassette for the expression of an
; TITLE OF INVENTION: endothiapsin precursor in Cryphonectria Parasitica
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P.O. BOX 299
; STREET: King Street Station, Suite 500, 1800 Diagonal Road
; CITY: ALEXANDRIA

RESULT 14
US-08-115-753-2
; Sequence 2, Application US/08115753
; Patent No. 6017762
; GENERAL INFORMATION:

APPLICANT: JARA, Patrick
APPLICANT: LEGOUX, Richard
APPLICANT: LOISON, Gerard
APPLICANT: RAZANAMPARANY, Voahangy
TITLE OF INVENTION: Cassette for the expression of an
endothiapepsin precursor in Cryptonectria Parasitica
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: P.O. BOX 299
STREET: King Street Station, Suite 500, 1800 Diagonal Road
CITY: ALEXANDRIA
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/115,753
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/758872
FILING DATE: 07-SEP-1991
APPLICATION NUMBER: FR 9011230
FILING DATE: 11-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, BERNHARD D
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/332
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 419 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
IMMEDIATE SOURCE:
CLONE: preproendothiapepsin
FEATURE:
NAME/KEY: Protein
LOCATION: 90..419
US-08-115-753-2

Query Match 8.3%; Score 155.5; DB 3; Length 419;
Best Local Similarity 25.1%; Pred. No. 7.4e-09;
Matches 82; Conservative 44; Mismatches 130; Indels 71; Gaps 17;
QY 5 IYLOAHT---LCSG--WSSTYRDLRKGVYVPYTOGKWEGLGTLVSIPIHGPNTVRAN 59
DB 147 IYTPSKSTAKLLSGATWSISYGD-----GSSSGDYITDTVSV--GGLTVTGOA- 194
QY 60 IAAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVKTHVFNLFSLQCCGAG 119
DB 195 VESAKVSSSFTEDSIDGLGLAFSTL-----NTVSTQQTFFDNKASLD 242
QY 120 FPLNSEVLASVGGSMIGIDHSLYTGLWTPYI--RREWYEVIIIVRVEINGDQDKMDC 178
DB 243 SPVFTADLGYHAPGYTFNGFIDTAYTGSIYTAVSTKOGFWETSTGYAVSGSTFKSTS 302
QY 179 KEYNYDKSLVDSGTNLRPKVFEAAVKSIRKAASSTER-----PPDGFWLGEQLVCQA 233
DB 303 ID-----GIADYGTLLLYLPATVVSAYVAQVSAKSSSSVGGYVFP-----C--S 345

QY 234 GTTPWNIPFVISLYLMGEVNTQSPRITILPOOYLRPVEDVATSQDDCYKFAISOSSTG-- 291
DB 346 ATLPSTFTGV-----GSARI-VIPGDYI-DFGPISGTSSSC--FGGIQSSAGIG 390
QY 292 -TVMGAVIMEGFYVVFDR--RKRIGFA 316
DB 391 INIFGDVALKAAFFVVGATTTLGF 417
RESULT 15
US-08-115-753-33
Sequence 33, Application US/08115753
Patent No. 6017762
GENERAL INFORMATION:
APPLICANT: JARA, Patrick
APPLICANT: LEGOUX, Richard
APPLICANT: LOISON, Gerard
APPLICANT: RAZANAMPARANY, Voahangy
TITLE OF INVENTION: Cassette for the expression of an
endothiapepsin precursor in Cryptonectria Parasitica
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: P.O. BOX 299
STREET: King Street Station, Suite 500, 1800 Diagonal Road
CITY: ALEXANDRIA
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/115,753
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/758872
FILING DATE: 07-SEP-1991
APPLICATION NUMBER: FR 9011230
FILING DATE: 11-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, BERNHARD D
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/332
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 419 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-115-753-33

Query Match 8.3%; Score 155.5; DB 3; Length 419;
Best Local Similarity 25.1%; Pred. No. 7.4e-09;
Matches 82; Conservative 44; Mismatches 130; Indels 71; Gaps 17;
QY 5 IYLOAHT---LCSG--WSSTYRDLRKGVYVPYTOGKWEGLGTLVSIPIHGPNTVRAN 59
DB 147 IYTPSKSTAKLLSGATWSISYGD-----GSSSGDYITDTVSV--GGLTVTGOA- 194
QY 60 IAAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVKTHVFNLFSLQCCGAG 119
DB 195 VESAKVSSSFTEDSIDGLGLAFSTL-----NTVSTQQTFFDNKASLD 242
QY 120 FPLNSEVLASVGGSMIGIDHSLYTGLWTPYI--RREWYEVIIIVRVEINGDQDKMDC 178

Db 243 SPVFTADLGYHAPGYNFGFIDTATTGTSITYTAVSTKQGFWEWTSTGYAVGSGTFKSTS 302
QY 179 KEYNDKSIVDSTTNLRPKKVFEEAAVKSIIKAASSTEK-----FPDGFWLGEOLVCHQA 233
Db 303 ID-----GIADGTTLTLLYPATVVSAYWQAQVSGAKSSSSVGGYVFP-----C--S 345
QY 234 GTPWNIFPVISLYLMGEVTNQSFRTIILPQOYLRPVEDVATSDDCYKFAISQSSTG-- 291
Db 346 ATLPSFTFGV-----GSARI-VIPGDYI-DFGPISGTSSSC--FGGIQSSAGIG 390
QY 292 -TVMGAVIMEGEYVVFDRA-RKRIGFA 316
Db 391 INIFGDVALKAAAFVVFENGATTPTLGFA 417

Search completed: March 24, 2001, 13:23:51
Job time: 141 sec

THIS PAGE BLANK (USPTO)